Update richtlijn Hoofd-halstumoren (6 modules)

Augustus 2021

INITIATIEF
Nederlandse Vereniging voor Keel-Neus-Oorheelkunde en Heelkunde van het Hoofd-Halsgebied

IN SAMENWERKING MET
Nederlandse Internisten Vereniging
Nederlandse Vereniging voor Mondziekten, Kaak- en Aangezichtschirurgie
Nederlandse Vereniging voor Nucleaire Geneeskunde
Nederlandse Vereniging voor Pathologie
Nederlandse Vereniging voor Plastische Chirurgie
Nederlandse Vereniging voor Radiologie
Nederlandse Vereniging voor Radiotherapie en Oncologie
Nederlandse Federatie van Kankerpatiëntenorganisaties | Patiëntenvereniging HOOFD-HALS
V&VN Oncologie

MET ONDERSTEUNING VAN
Kennisinstituut van de Federatie Medisch Specialisten

FINANCIERING
De richtlijnontwikkeling werd gefinancierd uit de Kwaliteitsgelden Medisch Specialisten (SKMS).

Colofon
UPDATE RICHTLIJN HOOFD-HALSTUMOREN (6 MODULES)

Nederlandse Vereniging voor Keel-Neus-Oorheelkunde en Heelkunde van het Hoofd-Halsgebied
Mercatorlaan 1200, 3528 BL Utrecht
Tel. 030 320 12 15
kno@kno.nl
https://www.kno.nl

Inhoudsopgave

Samenstelling van de werkgroep 4
Ontwikkelde modules (tot januari 2021) 6
Verantwoording 7
Module 1.1 Diagnostiek invasiediepte mondholtecarcinoom 16
Module 1.2 Diagnostiek botinvasie mandibula 76
Module 2.1 HPV-statusbepaling 126
Module 10.1 Aanvraag en verslag pathologie-onderzoek 196
Module 12.1 Behandeling T1-2N0-1 orofarynxcarcinomen 211
Module 12.4 Behandeling HPV-positieve orofarynx tumoren 252
Bijlage 1 Kennislacunes 274

Samenstelling van de werkgroep

Werkgroep

- Prof. Dr. R. de Bree, KNO-arts/hoofd-halschirurg, UMC Utrecht, Utrecht, NVKNO (voorzitter)
- Dr. M.B. Karakullukcu, KNO-arts/hoofd-halschirurg, NKI, Amsterdam, NVKNO
- Dr. H.P. Verschuur, KNO-arts/hoofd-halschirurg, Haaglanden MC, Den Haag, NVKNO
- Dr. M. Walenkamp, AIOS-KNO, LUMC, Leiden, NVKNO
- Dr. A. Sewnaik, KNO-arts/hoofd-halschirurg, Erasmus MC, Rotterdam, NVKNO
- Drs. L.H.E. Karssemakers, MKA-chirurg-oncoloog/hoofd-hals chirurg, NKI, Amsterdam, NVMKA
- Dr. M.J.H. Witjes, MKA-chirurg-oncoloog, UMC Groningen, Groningen, NVMKA
- Drs. L.A.A. Vaassen, MKA-chirurg-oncoloog, Maastricht UMC+, Maastricht, NVMKA
- Drs. W.L.J. Weijs, MKA-chirurg-oncoloog, Radboud UMC, Nijmegen, NVKMA
- Drs. E.M. Zwijnenburg, Radiotherapeut-oncoloog, Radboud UMC, Nijmegen, NVRO
- Dr. A. Al-Mamgani, Radiotherapeut-oncoloog, NKI, Amsterdam, NVRO
- Prof. Dr. C.H.J. Terhaard, Radiotherapeut-oncoloog, UMC Utrecht, Utrecht, NVRO
- Drs. J.G.M. Van den Hoek, Radiotherapeut-oncoloog, UMC Groningen, Groningen, NVRO
- Dr. E. Van Meerten, Internist-oncoloog, Erasmus MC Kanker Instituut, Rotterdam, NIV
- Dr. M. Slingerland, Internist-oncoloog, LUMC, Leiden, NIV
- Drs. M.A. Huijing, Plastisch Chirurg, UMC Groningen, Groningen, NVPC
- Prof. Dr. S.M. Willems, Klinisch patholoog, UMC Groningen, Groningen, NVVP
- Prof. Dr. E. Bloemena, Klinisch patholoog, Amsterdam UMC, locatie Vumc, Amsterdam, NVVP
- R.A. Burdorf, Voorzitter dagelijks bestuur patiëntenvereniging, Patiëntenvereniging HOOFD-HALS, PvHH
- P.S. Verdouw, Hoofd infocentrum patiëntenvereniging, Patiëntenvereniging HOOFD-HALS, PvHH
- A.A.M. Goossens, Verpleegkundig specialist oncologie, Haaglanden MC, Den Haag, V&VN
- Dr. P. de Graaf, Radioloog, Amsterdam UMC, Amsterdam, NVvR
- Dr. W.V. Vogel, Nucleair geneeskundige/radiotherapeut-oncoloog, NKI, Amsterdam, NVNG
- Drs. G.J.C. Zwezerijnen, Nucleair geneeskundige, Amsterdam UMC, Amsterdam, NVNG

<u>Klankbordgroep</u>

- Dr. C.M. Speksnijder, Fysiotherapeut, UMC Utrecht, Utrecht, KNGF
- Ir. A. Kok, Diëtist, UMC Utrecht, Utrecht, NVD
- Dr. M.M. Hakkesteegt, Logopedist, Erasmus MC, Rotterdam, NVLF
- Drs. D.J.M. Buurman, Tandarts-MFP, Maastricht UMC+, Maastricht, KNMT
- W. Van der Groot-Roggen, Mondhygiënist, UMC Groningen, Groningen, NVvM
- Drs. D.J.S. Dona, Bedrijfsarts/Klinisch arbeidsgeneeskundige oncologie, Radboud UMC, Nijmegen, NVKA
- Dr. M. Sloots, Ergotherapeut, UMC Utrecht, Utrecht
- J. Poelstra, Medisch maatschappelijk werkster, op persoonlijke titel

Met ondersteuning van

- Dr. J. Boschman, Senior adviseur, Kennisinstituut van de Federatie Medisch Specialisten
- Dr. C. Gaasterland, Adviseur, Kennisinstituut van de Federatie Medisch Specialisten
- Dr. A. Van der Hout, Adviseur, Kennisinstituut van de Federatie Medisch Specialisten
- Dr. L. Oostendorp, Adviseur, Kennisinstituut van de Federatie Medisch Specialisten
- Drs. M. Oerbekke, Adviseur, Kennisinstituut van de Federatie Medisch Specialisten
- Drs. A. Hoeven, Junior adviseur, Kennisinstituut van de Federatie Medisch Specialisten

Ontwikkelde modules (tot januari 2021)

In dit document kunt u de nieuwe/geüpdate richtlijnmodules vinden die binnen de Richtlijn Hoofdhals-tumoren worden ingebed.

Hoofdstuk	**Moduletitel**	**Vorm**	**Locatie in Richtlijn 2014 document**
Diagnostiek mondholte	Diagnostiek invasiediepte mondholtecarcinoom	Nieuw	-
Diagnostiek mondholte	Diagnostiek botinvasie mandibula	Update bestaande module (wordt een aparte module in de geüpdate richtlijn)	Pagina 7
Diagnostiek orofarynxcarcinoom	HPV-statusbepaling	Nieuw	-
Aanvraag en verslag pathologie	Aanvraag en verslag pathologie-onderzoek	Update bestaande module	Pagina 20
Behandeling orofarynxcarcinoom	Behandeling T1-2N0-1 orofarynxcarcinomen	Update bestaande module	Pagina 37
Behandeling orofarynxcarcinoom	Behandeling HPV-positieve tumoren	Nieuw	-

Verantwoording

Leeswijzer
De verantwoording wordt op de Richtlijnendatabase bij elke nieuwe of geüpdate module opgenomen. Aangezien deze richtlijn gedeeltelijk een herziening betreft, zal het gedeelte 'Autorisatie en geldigheid' in de richtlijn per module (kunnen) verschillen. Het overige gedeelte van de verantwoording is gelijk voor alle herziende of nieuwe modules, en wordt slechts éénmaal bijgevoegd.

Autorisatie en geldigheid
De geldigheid van de richtlijnmodule komt te vervallen indien nieuwe ontwikkelingen aanleiding zijn een herzieningstraject te starten.
NB: Informatie over de autorisatiedatum, autoriserende partij(en), herbevestiging en regiehouder(s) worden te zijner tijd na autorisatie toegevoegd aan deze alinea.

Algemene gegevens
De ontwikkeling/herziening van deze richtlijnmodule werd ondersteund door het Kennisinstituut van de Federatie Medisch Specialisten (www.demedischspecialist.nl/kennisinstituut) en werd gefinancierd uit de Stichting Kwaliteitsgelden Medisch Specialisten (SKMS). De financier heeft geen enkele invloed gehad op de inhoud van de richtlijnmodule.

Samenstelling werkgroep
Voor het ontwikkelen van de richtlijnmodule is in 2019 een multidisciplinaire werkgroep ingesteld, bestaande uit vertegenwoordigers van alle relevante specialismen (zie hiervoor de Samenstelling van de werkgroep) die betrokken zijn bij de zorg voor patiënten met hoofd-halstumoren.

Belangenverklaringen
De Code ter voorkoming van oneigenlijke beïnvloeding door belangenverstrengeling is gevolgd. Alle werkgroepleden hebben schriftelijk verklaard of zij in de laatste drie jaar directe financiële belangen (betrekking bij een commercieel bedrijf, persoonlijke financiële belangen, onderzoeksfinanciering) of indirecte belangen (persoonlijke relaties, reputatiemanagement) hebben gehad. Gedurende de ontwikkeling of herziening van een module worden wijzigingen in belangen aan de voorzitter doorgegeven. De belangenverklaring wordt opnieuw bevestigd tijdens de commentaarfase.

Een overzicht van de belangen van werkgroepleden en het oordeel over het omgaan met eventuele belangen vindt u in onderstaande tabel. De ondertekende belangenverklaringen zijn op te vragen bij het secretariaat van het Kennisinstituut van de Federatie Medisch Specialisten.

Werkgroepl id	Functie	Nevenfuncties	Gemelde belangen	Ondernomen actie
Bree, de	KNO-arts/hoofd-halschirurg, UMC Utrecht	* Lid Algemeen Bestuur Patiëntenvereniging Hoofd-Hals (onbetaald) * Voorzitter Research Stuurgroep NWHHT * Lid Richtlijnen commissie NWHHT	Geen	Geen

		* Lid dagelijks bestuur NWHHT * Lid Clinical Audit Board van de Dutch Head and Neck Audit (DHNA) * Lid wetenschappelijk adviescommissie DORP * Voorzitter Adviescommissie onderzoek hoofd-halskanker (IKNL/PALGA/DHNA/NWHHT)		
Slingerland	Internist-oncoloog, LUMC	* 2018-present: Treasurer of the "Dutch Association of Medical Oncology"(NVMO - vacancy fees) * 2018-present: Member of the "Dutch Working Group for Head-Neck Tumors" (NWHHT-Systemic therapy) * 2016-present: Member of the 'Dutch Working Group for Head-Neck Tumors" (NWHHT - study group steering group (coordinating)) * 2016-present: Member of the "Dutch Working Group for Head-Neck Tumors" (NWHHT - Elderly Platform) * 2012-present: Member "Working Group for Head-Neck Tumors" (WHHT) "University Cancer Centre"(UCK) Leiden - Den Haag * 2019: Member CAB DHNA	Deelname Nationaal expert forum hoofd-halskanker MSD dd 2-5-2018 * Deelname Checkmate studie, sponsor Bristol-Myers Squibb (BMS): An open label, randomized phase 3 clinical trial of nivolumab versus therapy of investigator's choice in recurrent or metastatic platinum-refractory squamous cell carcinoma of the head and neck (SCCHN) * Deelname Commence studie, sponsor Radboud University, in collaboration with Merck Serono International SA (among several Dutch medical centers): A phase lB-II study of the combination of cetuximab and methotrexate in recurrent of metastatic squamous cell carcinoma of the head and neck. A study of the Dutch Head and Neck Society, MOHN01/COMMENCE study. * Deelname HESPECTA studie: Phase I study: to determine the biological activity of two HPV16E6 specific peptides coupled to Amplivant®, a Toll-like receptor ligand in non-metastatic patients treated for HPV16-positive head and neck cancer. * Deelname PINCH studie (nog niet open): PD-L1 ImagiNg to predict durvalumab treatment response in HNSCC (PINCH) trial; patiënten met biopt bewezen locally recurrent of gemetastaseerd HNSCC * Deelname ISA 101b-HN-01-17 studie (nog niet open): A randomized, Double-blind, Placebo-Controlled, Phase 2 Study of Cemiplimab versus the combination of Cemiplimab with	In de werkgroep participeren 2 internist-oncologen, zodat één van beide de voortrekker is van modules over systemische therapie. Actie: werkgroeplid is uitgesloten van besluitvorming bij modules die betrekking hebben op de onderwerpen van de gemelde onderzoeken: nivolumab, cetuximab + methotrexaat, Amplivant, durvalumab, cemiplimab.

			ISA101b in the Treatment of Subjects.	
Meerten, van	Internist-oncoloog, Erasmus MC Kanker Instituut	Geen	Op dit moment Principal Investigator voor NL van gerandomiseerde fase III trial naar toegevoegde waarde van pembrolizumab aan chemoradiotherapie bij patiënten met gevorderd hoofdhalskanker. Sponsor: GlaxoSmithKline Research & Development Ltd. Studie is nog lopend, resultaten zullen pas bekend zijn na verschijning van de richtlijn. In toekomst mogelijk participatie aan door industrie gesponsorde studies op gebied van behandeling van hoofdhalskanker	In de werkgroep participeren 2 internist-oncologen, zodat één van beide de voortrekker is van modules over systemische therapie. Actie: werkgroeplid is uitgesloten van besluitvorming bij modules die betrekking hebben op het onderwerp van het gemelde onderzoeken: de toegevoegde waarde van pembrolizumab bij patiënten met gevorderd hoofdhalskanker.
Huijing	Plastisch chirurg, UMC Groningen	Geen	Geen	Geen
Sewnaik	KNO-arts/hoofd Hals chirurg, Erasmus MC	Sectorhoofd Hoofd-Hals chirurgie	Geen	Geen
Vaassen	MKA-chirurg-oncoloog, Maastricht UMC+ / CBT Zuid-Limburg	*Lid Bestuur NVMKA *Waarnemend hoofd MKA-chirurgie MUMC	Geen	Geen
Witjes	MKA-chirurg-oncoloog, UMC Groningen	Geen	PI van KWF grant: RUG 2015 - 8084: Image guided surgery for margin assessment of head & neck Cancer using cetuximab-IRDye800 cONjugate (ICON) geen financieel belang	Geen. Financiering door KWF werd niet als een belang ingeschat.
Bloemena	Klinisch patholoog, Amsterdam UMC (locatie Vumc) / Radboud UMC / Academisch Centrum voor Tandheelkunde Amsterdam (ACTA)	* Lid bestuur Nederlandse Vereniging voor Pathologie (NVVP) – vacatiegeld (tot 1-12-20) * Voorzitter Commissie Bij- en Nascholing (NVVP) * Voorzitter (tot 1-12-20) Wetenschappelijke Raad PALGA - onbezoldigd	Geen	Geen

Willems	Klinisch patholoog, UMC Groningen	Vice-vz PALGA, AB NWHHT, CAB DHNA, mede-vz en oprichter expertisegroep HH pathologie NL, Hoofdhalspathologie UMC Groningen	PDL1 trainer NL voor MSD Onderzoeksfinanciering van Pfizer, Roche, MSD, BMS, Lilly, Novartis, Bayer, Amge, AstraZeneca	Geen
Karakullukcu	KNO-arts/hoofd-hals chirurg, NKI/AVL	Geen	Geen	Geen
Verschuur	KNO-arts/Hoofd-hals chirurg, Haaglanden MC	* Opleider KNO-artsen * Dagvoorzitter	Geen	Geen
Walenkamp	AIOS KNO, LUMC	Geen	Geen	Geen
Al-Mamgani	Radiotherapeut-oncoloog, NKI/AVL	Geen	Geen	Geen
Terhaard	Radiotherapeut-oncoloog, UMC Utrecht	Niet van toepassing	Geen	Geen
Hoek, van den	Radiotherapeut-oncoloog UMCG	Niet van toepassing	Geen	Geen
Zwijnenburg	Radiotherapeut, Hoofd-hals Radboud UMC	Geen	Geen	Geen
Burdorf	Patiëntvertegenwoordiger	Geen	Geen	Geen
Verdouw	Hoofd Infocentrum patiëntenvereniging HOOFD HALS	Geen	Werkzaam bij de patiëntenvereniging. De achterban heeft baat bij een herziening van de richtlijn	Geen
Karssemakers	Hoofd-hals chirurg NKI/AVL MKA-chirurg-oncoloog Amsterdam UMC (locatie AMC) / vakgroep kaakchirurgie Amsterdam West	Niet van toepassing	Geen	Geen
Goossens	Verpleegkundig specialist, Haaglanden Medisch Centrum (HMC)	* Bestuurslid (penningmeester) PWHHT (onbetaald) * Lid Commissie voorlichting PVHH (onbetaald)	Geen	Geen
Zwezerijnen	Nucleair geneeskundige, Amsterdam UMC (locatie Vumc) PhD kandidaat, Amsterdam UMC (locatie Vumc)	Lid als nucleair geneeskundige in HOVON imaging werkgroep (bespreken van richtlijnen en opzetten/uitvoeren van wetenschappelijke studies met betrekking tot beeldvorming in de hematologie); onbetaald	Geen	Geen
Vogel	Nucleair geneeskundige/ra	Geen	In de afgelopen jaren incidenteel advies of onderwijs, betaald	Geen

	diotherapeut-oncoloog, AVL		door Bayer, maar niet gerelateerd aan hoofd-hals KWF-grant speekselklier toxiteit na behandeling. Geen belang bij de richtlijn	
Graaf, de	Radioloog, Amsterdam UMC (locatie Vumc)	Bestuurslid sectie Hoofd-Hals radiologie (onbetaald)	Geen	Geen
Weijs	MKA-chirurg-oncoloog, Radboudumc	MKA-chirurg, Weijsheidstand B.V. Werkzaam als algemeen praktiserend MKA-chirurg, betaald (0,1 fte)	Geen	Geen

Inbreng patiëntenperspectief
Er werd aandacht besteed aan het patiëntenperspectief door het uitnodigen van de patiëntenvereniging HOOFD-HALS (PVHH) voor de Invitational conference en met afgevaardigden van de PVHH in de werkgroep. Het verslag hiervan (zie aanverwante producten) is besproken in de werkgroep. De verkregen input is meegenomen bij het opstellen van de uitgangsvragen, de keuze voor de uitkomstmaten en bij het opstellen van de overwegingen. De conceptrichtlijn is tevens voor commentaar voorgelegd aan de patiëntenvereniging HOOFD-HALS en de eventueel aangeleverde commentaren zijn bekeken en verwerkt.

Werkwijze
AGREE
Deze richtlijnmodule is opgesteld conform de eisen vermeld in het rapport Medisch Specialistische Richtlijnen 2.0 van de adviescommissie Richtlijnen van de Raad Kwaliteit. Dit rapport is gebaseerd op het AGREE II instrument (Appraisal of Guidelines for Research & Evaluation II; Brouwers, 2010).

Knelpuntenanalyse en uitgangsvragen
Tijdens de voorbereidende fase inventariseerden de werkgroep de knelpunten in de zorg voor patiënten met hoofd-halstumoren. De werkgroep beoordeelde de aanbeveling(en) uit de eerdere richtlijnmodule (NVKNO, 2014) op noodzaak tot revisie. Tevens zijn er knelpunten aangedragen door de patiëntenvereniging en genodigde partijen tijdens de invitational conference (zie aanverwante producten voor het verslag van de invitational conference). Op basis van de uitkomsten van de knelpuntenanalyse zijn door de werkgroep concept-uitgangsvragen opgesteld en definitief vastgesteld.

Uitkomstmaten
Na het opstellen van de zoekvraag behorende bij de uitgangsvraag inventariseerde de werkgroep welke uitkomstmaten voor de patiënt relevant zijn, waarbij zowel naar gewenste als ongewenste effecten werd gekeken. Hierbij werd een maximum van acht uitkomstmaten gehanteerd. De werkgroep waardeerde deze uitkomstmaten volgens hun relatieve belang bij de besluitvorming rondom aanbevelingen, als cruciaal (kritiek voor de besluitvorming), belangrijk (maar niet cruciaal) en onbelangrijk. Tevens definieerde de werkgroep tenminste voor de cruciale uitkomstmaten welke verschillen zij klinisch (patiënt) relevant vonden.

Methode literatuursamenvatting
Een uitgebreide beschrijving van de strategie voor zoeken en selecteren van literatuur en de beoordeling van de risk-of-bias van de individuele studies is te vinden onder 'Zoeken en selecteren' onder Onderbouwing. De beoordeling van de kracht van het wetenschappelijke bewijs wordt hieronder toegelicht.

Beoordelen van de kracht van het wetenschappelijke bewijs
De kracht van het wetenschappelijke bewijs werd bepaald volgens de GRADE-methode. GRADE staat voor 'Grading Recommendations Assessment, Development and Evaluation' (zie http://www.gradeworkinggroup.org/). De basisprincipes van de GRADE-methodiek zijn: het benoemen en prioriteren van de klinisch (patiënt) relevante uitkomstmaten, een systematische review per uitkomstmaat, en een beoordeling van de bewijskracht per uitkomstmaat op basis van de acht GRADE-domeinen (domeinen voor downgraden: risk of bias, inconsistentie, indirectheid, imprecisie, en publicatiebias; domeinen voor upgraden: dosis-effect relatie, groot effect, en residuele plausibele confounding).

GRADE onderscheidt vier gradaties voor de kwaliteit van het wetenschappelijk bewijs: hoog, redelijk, laag en zeer laag. Deze gradaties verwijzen naar de mate van zekerheid die er bestaat over de literatuurconclusie, in het bijzonder de mate van zekerheid dat de literatuurconclusie de aanbeveling adequaat ondersteunt (Schünemann, 2013; Hultcrantz, 2017).

GRADE	Definitie
Hoog	• er is hoge zekerheid dat het ware effect van behandeling dicht bij het geschatte effect van behandeling ligt; • het is zeer onwaarschijnlijk dat de literatuurconclusie klinisch relevant verandert wanneer er resultaten van nieuw grootschalig onderzoek aan de literatuuranalyse worden toegevoegd.
Redelijk	• er is redelijke zekerheid dat het ware effect van behandeling dicht bij het geschatte effect van behandeling ligt; • het is mogelijk dat de conclusie klinisch relevant verandert wanneer er resultaten van nieuw grootschalig onderzoek aan de literatuuranalyse worden toegevoegd.
Laag	• er is lage zekerheid dat het ware effect van behandeling dicht bij het geschatte effect van behandeling ligt; • er is een reële kans dat de conclusie klinisch relevant verandert wanneer er resultaten van nieuw grootschalig onderzoek aan de literatuuranalyse worden toegevoegd.
Zeer laag	• er is zeer lage zekerheid dat het ware effect van behandeling dicht bij het geschatte effect van behandeling ligt; • de literatuurconclusie is zeer onzeker.

Bij het beoordelen (graderen) van de kracht van het wetenschappelijk bewijs in richtlijnen volgens de GRADE-methodiek spelen grenzen voor klinische besluitvorming een belangrijke rol (Hultcrantz, 2017). Dit zijn de grenzen die bij overschrijding aanleiding zouden geven tot een aanpassing van de aanbeveling. Om de grenzen voor klinische besluitvorming te bepalen moeten alle relevante uitkomstmaten en overwegingen worden meegewogen. De grenzen voor klinische besluitvorming zijn daarmee niet één op één vergelijkbaar met het minimaal klinisch relevant verschil (Minimal Clinically Important Difference, MCID). Met name in situaties waarin een interventie geen belangrijke nadelen heeft en de kosten relatief laag zijn, kan de grens voor klinische besluitvorming met betrekking tot de effectiviteit van de interventie bij een lagere waarde (dichter bij het nuleffect) liggen dan de MCID (Hultcrantz, 2017).

Overwegingen (van bewijs naar aanbeveling)
Om te komen tot een aanbeveling zijn naast (de kwaliteit van) het wetenschappelijke bewijs ook andere aspecten belangrijk en worden meegewogen, zoals aanvullende argumenten uit

bijvoorbeeld de biomechanica of fysiologie, waarden en voorkeuren van patiënten, kosten (middelenbeslag), aanvaardbaarheid, haalbaarheid en implementatie. Deze aspecten zijn systematisch vermeld en beoordeeld (gewogen) onder het kopje 'Overwegingen' en kunnen (mede) gebaseerd zijn op expert opinion. Hierbij is gebruik gemaakt van een gestructureerd format gebaseerd op het evidence-to-decision framework van de internationale GRADE Working Group (Alonso-Coello, 2016a; Alonso-Coello, 2016b). Dit evidence-to-decision framework is een integraal onderdeel van de GRADE-methodiek.

Formuleren van aanbevelingen

De aanbevelingen geven antwoord op de uitgangsvraag en zijn gebaseerd op het beschikbare wetenschappelijke bewijs en de belangrijkste overwegingen, en een weging van de gunstige en ongunstige effecten van de relevante interventies. De kracht van het wetenschappelijk bewijs en het gewicht dat door de werkgroep wordt toegekend aan de overwegingen, bepalen samen de sterkte van de aanbeveling. Conform de GRADE-methodiek sluit een lage bewijskracht van conclusies in de systematische literatuuranalyse een sterke aanbeveling niet a priori uit, en zijn bij een hoge bewijskracht ook zwakke aanbevelingen mogelijk (Agoritsas, 2017; Neumann, 2016). De sterkte van de aanbeveling wordt altijd bepaald door weging van alle relevante argumenten tezamen. De werkgroep heeft bij elke aanbeveling opgenomen hoe zij tot de richting en sterkte van de aanbeveling zijn gekomen.

In de GRADE-methodiek wordt onderscheid gemaakt tussen sterke en zwakke (of conditionele) aanbevelingen. De sterkte van een aanbeveling verwijst naar de mate van zekerheid dat de voordelen van de interventie opwegen tegen de nadelen (of vice versa), gezien over het hele spectrum van patiënten waarvoor de aanbeveling is bedoeld. De sterkte van een aanbeveling heeft duidelijke implicaties voor patiënten, behandelaars en beleidsmakers (zie onderstaande tabel). Een aanbeveling is geen dictaat, zelfs een sterke aanbeveling gebaseerd op bewijs van hoge kwaliteit (GRADE-gradering HOOG) zal niet altijd van toepassing zijn, onder alle mogelijke omstandigheden en voor elke individuele patiënt.

Implicaties van sterke en zwakke aanbevelingen voor verschillende richtlijngebruikers		
	Sterke aanbeveling	*Zwakke (conditionele) aanbeveling*
Voor patiënten	De meeste patiënten zouden de aanbevolen interventie of aanpak kiezen en slechts een klein aantal niet.	Een aanzienlijk deel van de patiënten zouden de aanbevolen interventie of aanpak kiezen, maar veel patiënten ook niet.
Voor behandelaars	De meeste patiënten zouden de aanbevolen interventie of aanpak moeten ontvangen.	Er zijn meerdere geschikte interventies of aanpakken. De patiënt moet worden ondersteund bij de keuze voor de interventie of aanpak die het beste aansluit bij zijn of haar waarden en voorkeuren.
Voor beleidsmakers	De aanbevolen interventie of aanpak kan worden gezien als standaardbeleid.	Beleidsbepaling vereist uitvoerige discussie met betrokkenheid van veel stakeholders. Er is een grotere kans op lokale beleidsverschillen.

Organisatie van zorg

In de knelpuntenanalyse en bij de ontwikkeling van de richtlijnmodule is expliciet aandacht geweest voor de organisatie van zorg: alle aspecten die randvoorwaardelijk zijn voor het verlenen van zorg (zoals coördinatie, communicatie, (financiële) middelen, mankracht en infrastructuur). Randvoorwaarden die relevant zijn voor het beantwoorden van deze specifieke uitgangsvraag zijn genoemd bij de overwegingen. Meer algemene, overkoepelende, of bijkomende aspecten van de organisatie van zorg worden behandeld in de module Organisatie van zorg.

Commentaar- en autorisatiefase

De conceptrichtlijnmodule werd aan de betrokken (wetenschappelijke) verenigingen en (patiënt) organisaties voorgelegd ter commentaar. De commentaren werden verzameld en besproken met de werkgroep. Naar aanleiding van de commentaren werd de conceptrichtlijnmodule aangepast en definitief vastgesteld door de werkgroep. De definitieve richtlijnmodule werd aan de deelnemende (wetenschappelijke) verenigingen en (patiënt) organisaties voorgelegd voor autorisatie en door hen geautoriseerd dan wel geaccordeerd.

Literatuur

Agoritsas T, Merglen A, Heen AF, Kristiansen A, Neumann I, Brito JP, Brignardello-Petersen R, Alexander PE, Rind DM, Vandvik PO, Guyatt GH. UpToDate adherence to GRADE criteria for strong recommendations: an analytical survey. BMJ Open. 2017 Nov 16;7(11):e018593. doi: 10.1136/bmjopen-2017-018593. PubMed PMID: 29150475; PubMed Central PMCID: PMC5701989.

Alonso-Coello P, Schünemann HJ, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, Treweek S, Mustafa RA, Rada G, Rosenbaum S, Morelli A, Guyatt GH, Oxman AD; GRADE Working Group. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. BMJ. 2016 Jun 28;353:i2016. doi: 10.1136/bmj.i2016. PubMed PMID: 27353417.

Alonso-Coello P, Oxman AD, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, Treweek S, Mustafa RA, Vandvik PO, Meerpohl J, Guyatt GH, Schünemann HJ; GRADE Working Group. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: Clinical practice guidelines. BMJ. 2016 Jun 30;353:i2089. doi: 10.1136/bmj.i2089. PubMed PMID: 27365494.

Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, Fervers B, Graham ID, Grimshaw J, Hanna SE, Littlejohns P, Makarski J, Zitzelsberger L; AGREE Next Steps Consortium. AGREE II: advancing guideline development, reporting and evaluation in health care. CMAJ. 2010 Dec 14;182(18):E839-42. doi: 10.1503/cmaj.090449. Epub 2010 Jul 5. Review. PubMed PMID: 20603348; PubMed Central PMCID: PMC3001530.

Hultcrantz M, Rind D, Akl EA, Treweek S, Mustafa RA, Iorio A, Alper BS, Meerpohl JJ, Murad MH, Ansari MT, Katikireddi SV, Östlund P, Tranæus S, Christensen R, Gartlehner G, Brozek J, Izcovich A, Schünemann H, Guyatt G. The GRADE Working Group clarifies the construct of certainty of evidence. J Clin Epidemiol. 2017 Jul;87:4-13. doi: 10.1016/j.jclinepi.2017.05.006. Epub 2017 May 18. PubMed PMID: 28529184; PubMed Central PMCID: PMC6542664.

Medisch Specialistische Richtlijnen 2.0 (2012). Adviescommissie Richtlijnen van de Raad Kwalitieit. http://richtlijnendatabase.nl/over_deze_site/over_richtlijnontwikkeling.html.

Neumann I, Santesso N, Akl EA, Rind DM, Vandvik PO, Alonso-Coello P, Agoritsas T, Mustafa RA, Alexander PE, Schünemann H, Guyatt GH. A guide for health professionals to interpret and use recommendations in guidelines developed with the GRADE approach. J Clin Epidemiol. 2016 Apr;72:45-55. doi: 10.1016/j.jclinepi.2015.11.017. Epub 2016 Jan 6. Review. PubMed PMID: 26772609.

Schünemann H, Brożek J, Guyatt G, et al. GRADE handbook for grading quality of evidence and strength of recommendations. Updated October 2013. The GRADE Working Group, 2013. Available from http://gdt.guidelinedevelopment.org/central_prod/_design/client/handbook/handbook.html.

Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, Williams JW Jr, Kunz R, Craig J, Montori VM, Bossuyt P, Guyatt GH; GRADE Working Group. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. BMJ. 2008 May 17;336(7653):1106-10. doi: 10.1136/bmj.39500.677199.AE. Erratum in: BMJ. 2008 May 24;336(7654). doi: 10.1136/bmj.a139.

Schünemann, A Holger J (corrected to Schünemann, Holger J). PubMed PMID: 18483053; PubMed Central PMCID: PMC2386626.

Wessels M, Hielkema L, van der Weijden T. How to identify existing literature on patients' knowledge, views, and values: the development of a validated search filter. J Med Libr Assoc. 2016 Oct;104(4):320-324. PubMed PMID: 27822157; PubMed Central PMCID: PMC5079497.

Module 1.1 Diagnostiek invasiediepte mondholtecarcinoom

Uitgangsvraag

Hoe zou de invasiediepte van mondholtecarcinomen bepaald dienen te worden?

Inleiding

Invasiediepte van de primaire tumor is een prognostische factor. In de nieuwe TNM-classificatie is de invasiediepte opgenomen als belangrijke parameter voor het stadiëren van mondholtecarcinomen. Ook is bij mondholtecarcinomen de invasiediepte een voorspeller voor de aanwezigheid van lymfekliermetastasen. De invasiediepte is de afstand van de (gereconstrueerde) mucosa tot het diepste punt van de tumor in het weefsel. Dit is niet gelijk aan de tumordikte. Bij ulceratieve tumoren is de tumordikte kleiner dan de invasiediepte, bij exofytisch groeiende tumoren is het omgekeerde het geval. Er worden diverse technieken gebruikt om preoperatief de invasiediepte te bepalen, maar het is nog onduidelijk wat de beste modaliteit is om mondholte carcinomen preoperatief te stadiëren.

Search and select

A systematic review of the literature was performed to answer the following question:

> What is the agreement between preoperative clinical examination (by palpation), computed tomography (CT), positron emission tomography/computed tomography (PET-CT), magnetic resonance imaging (MRI) or intraoral ultrasound, and postoperative histopathologic results for measuring the depth of the invasion (or tumor thickness) by a tumor in patients with an oral cavity carcinoma?

P: patients with an oral cavity carcinoma;
I: preoperative determination of the depth of invasion (or tumor thickness) with palpation, CT, PET-CT, MRI, or intraoral ultrasound;
C: comparisons between palpation, CT, PET-CT, MRI, or intraoral ultrasound with postoperative pathological assessment as a reference standard;
O: agreement parameters on a continuous (depth in millimeters) or categorical (at a threshold, or for T-stage) measurement level.

<u>Relevant outcome measures</u>

The guideline development group considered agreement parameters regarding the final T-staging of the tumor and agreement on a continuous measurement level as a critical outcome measures for decision making.

A priori, the working group did not define the outcome measures listed above but used the definitions used in the studies.

The working group defined an underestimation and overestimation of >2 millimeter compared to the postoperative pathological assessment as a clinically important disagreement. This is acknowledged to be an arbitrary choice, since evidence regarding the clinical importance of the 2-millimeter border is lacking. A Cohen's Kappa (K) was considered sufficient when the K was greater or equal to 0.70 (Terwee, 2007; Prinssen, 2016).

The working group defined the time between preoperative assessment and the surgical resection shorter than or equal to 4 weeks as adequate. This is acknowledged to be an arbitrary interval, however it was presumed that this period would usually not allow a change in the construct to be measured.

Search and select (Methods)
The databases Medline (via OVID) and Embase (via Embase.com) were searched with relevant search terms until 12th of November 2019 for systematic reviews and primary diagnostic studies. The detailed search strategy is depicted under the tab Methods. The systematic literature search resulted in 311 hits. Studies were selected based on the following criteria: patients had an oral cavity carcinoma, agreement between preoperative assessment of the depth of invasion or tumor thickness with palpation (clinical examination) /CT/PET-CT/MRI/intraoral ultrasound and a postoperative pathological assessment was reported, reported parameters were for absolute agreement or these could be calculated. Initially 35 studies were selected after the screening of title and abstract. The working group checked the methods of the full-text studies to determine whether 'depth of invasion' or 'tumor thickness' was measured. After reading the full text, 24 studies were excluded (see the table with reasons for exclusion under the tab Evidence tables). Ten primary studies and one systematic review were included.

Results
Six primary studies were included in the analyses of literature for depth of invasion. One systematic review (10 studies provided information) and four primary studies were included for tumor thickness. Important study characteristics and results were extracted in the evidence tables. Results are summarized in Table 1.1 (depth of invasion) and Table 1.2 (tumor thickness) under the 'summary of literature'. The assessment of the risk of bias is summarized in the risk of bias tables (under the tab Evidence tables).

Risk of bias was assessed with the *Consensus-based Standards for the selection of health Measurement INstruments* (COSMIN) risk of bias checklist (Mokkink, 2010). The boxes concerning reliability and measurement error were used for the risk of bias assessment, since the clinical question concerned inter-instrument reliability and/or inter-instrument agreement. The conclusive risk of bias outcome is the lowest score on the COSMIN 4-point risk of bias tool (i.e. the lowest-score-counts principle). The study design and procedures for each assessed instrument was assessed. For example, when a study assessed both CT and MRI measurements (separately) versus histopathological measurements, both the design and procedures of CT versus histopathology and MRI versus histopathology are assessed individually for potential risk of bias. A preoperative assessment with an interval of 4 weeks or shorter before surgery was deemed appropriate.

The adapted GRADE assessment was conducted in accordance with the described procedure by Mokkink (2018). The adapted GRADE procedure entailed that three levels could be downgraded in the risk of bias domain: one level for a serious risk (multiple studies of doubtful quality or one study of adequate quality), two levels for a very serious risk (multiple studies of inadequate quality or one study of doubtful quality), or three levels for an extremely serious risk (only one study of inadequate quality). The inconsistency domain could be downgraded by one or two levels when there was unexplained heterogeneity between the reported outcomes. A maximum of two levels could be downgraded for imprecision: one level (body of evidence contains n=50 to n=100), or two levels (body of evidence contains less than n=50). When the included study did not completely match the PICO as defined in this guideline module, one or two levels could be downgraded for indirectness. Publication bias is not assessed in the adjusted GRADE procedure.

Summary of literature

Description of studies included for the agreement on depth of invasion

Alsaffar (2016) assessed the agreement between palpation or MR images and histopathology for the depth of invasion. The study recruited patients with newly diagnosed oral squamous cell carcinoma (n=53) of which there were 34 males. The mean age was 64 (SD or range not reported). Various T-stages (T1: n=22, T2: n=22, T3: n=7, T4: n=2) and N-stages (N0: n=32, N1: n=7, N2: n=11) were in the sample. It was unclear which staging system was used, however it is likely the AJCC TNM-staging system (presumably the 7th edition) was used. The palpation was performed by the treating surgeon, prior to the radiological assessment. Preoperative MRI was performed and the depth of invasion was measured from the adjacent mucosa to the deepest tumor invasion. The time period between the preoperative assessments and the histopathological assessment (on formalin fixed specimens) was unclear. Tumor invasion was categorized in two categories: < 5 millimeters and ≥ 5 millimeters. A Cohen's Kappa was calculated to assess the agreement.

Goel (2016) recruited patients (n=61) to assess the agreement between clinical examination or MRI and histopathology for the depth of invasion (categorized in T-stages) in patients with biopsy proven squamous cell carcinomas of the tongue or gingiva-buccal area. Forty-five of the included patients were male and various T-stages were in the sample (T1: n=4, T2: n=16, T3: n=13, T4: n=28). A TNM staging system was used (unclear edition). No procedures were described for the clinical examination or histopathological assessment. However, the tissue was probably fixed with formalin. MRI was performed with a 1.5T scanner (used sequence: axial and coronal T2WI, postcontrast T1WI). The time period between the preoperative assessments and histopathological assessment was unclear. Agreement between the clinical examination or MR imaging and histopathology on the T-stage was calculated with a Cohen's Kappa.

Iida (2018) assessed the agreement on depth of invasion between ultrasound and histopathology in patients with an early oral tongue squamous cell carcinoma between June 2008 and December 2015. Fifty-six patients were included, with a mean age of 59 years (range: 25 to 90) and of which 34 were male. All participants had their carcinoma located on the lateral ledge of the tongue. Tumor stage was not reported for the participants. It was unclear if and which edition of a staging system was used. The ultrasound assessment was performed in an outpatient clinic, using a 16-MHz scanner and a T-shaped ultrasonographic probe, where the patient extended their tongues during the preoperative ultrasound measurement. Histopathological assessment was performed with a micrometer in the tumor specimen, which was formalin-fixed and paraffin-embedded. The time period between preoperative ultrasound and postoperative histopathology was unclear. The depth of invasion was categorized by a threshold, resulting in two categories: < 5 millimeters and ≥ 5 millimeters. A Cohen's Kappa was calculated for the agreement.

Mao (2019) investigated the agreement between MR imaging and histopathology in patients first diagnosed with squamous cell carcinoma of the tongue (n=150). The mean age of patients was 58 years (SD: 12.1). There were 80 males and 70 females in the sample, with various tumor locations: ventral side of the tongue (n=35), border of the tongue (n=89), dorsal side of the tongue (n=19), and the base of the tongue (n=7). Several tumor morphologies were identified: ulcer type (n=41), invasive type (n=94), and exogeneous type (n=15). Participants had a T-stage of T1 (n=43), T2 (n=71), or T3 (n=36) and an N-stage of N1 (n=16), N2b (n=17), or N2c (n=2). The 7th edition of the AJCC staging system was used. A 1.5T MR scanner was used 1 week preoperatively to measure the depth of invasion with a section thickness of 1 millimeter (used sequences: T1 axial, coronal and sagittal sequences, T2 axial

and coronal sequences with fat suppression, T1-weighted axial, coronal and sagittal sequences with fat suppression and contrast media). Surgical tumor specimens were preserved in formalin. Pathological sections and staining were performed to measure the tumor invasion. Agreement was quantified in Bland-Altman plots for all participants, per T-stage, and per tumor morphology.

Verma (2019) assessed the agreement between MR imaging and histopathology for tumor thickness (per T-stage) in patients with biopsy proven squamous cell carcinoma of the tongue (n=50). The sample consisted of 38 males and 12 females with mean age of the sample was 49 (SD not reported). Various T-stages were prevalent in the sample: T1 (n=24), T2 (n=18), and T3 (n=8). The 7th and 8th edition of the AJCC staging system were used for the study. No other characteristics were reported. Tumor thickness was preoperatively assessed with MR imaging (4 millimeter slices, used sequences: T1W1 axial and coronal, T2WI axial, coronal and saggital, coronal STIR, and postcontrast axial T1W). Tumor dimensions (anteroposteriorly, mediolaterally, superoinferiorly) were measured. Tumor thickness was measured in three dimensions with histopathology (presumably on formalin fixed material), however no further procedures were reported. The time period between the preoperative MR imaging and the histopathological assessment was unclear. A Cohen's Kappa was not calculated by the authors, but could be calculated from the presented 3-by-3 table showing the T-classifications of MR imaging and histopathology.

Vidiri (2019) assessed the depth of invasion as well as tumor thickness (per T-stage) in patients diagnosed with oral tongue squamous cell carcinoma between 2013 and 2018. The median age for the patients (n=43, 18 males and 25 females) was 65 with a range from 31 to 81. Various T-stages were prevalent: T1 (n=10), T2 (n=12), and T3 (n=21). The 8th edition of the AJCC staging system was used. Preoperative MR imaging was performed with a 1.5T scanner 3 to 4 weeks preoperatively (used sequences: coronal T2W, axial FSE T2W, pre-contrast axial T1WI, DWI through single-shot spin-echo and echo-planar imaging). Two radiologists, one experienced and one inexperienced, assessed the images independently from each other. Agreement between histopathology (on formalin fixed material) and the results of both radiologists were reported separately. Resected tissue was fixed in formalin. Embedding, sectioning, and staining (with hematoxylin and eosin) was performed for histopathological analyses. Bland-Altman plots were reported for the depth of invasion and Cohen's Kappa was reported for agreement on T-stage as tumor thickness.

Studies included for the agreement on tumor thickness

Brouwer de Koning (2019) investigated the agreement between ultrasound or MR imaging and histopathology for tumor thickness in clinically stages T1-2 oral cavity carcinomas. MR images were acquired between 2011 and 2016. A total of 83 patients were included in the analyses with a mean age of 61 years (range: 31 to 88). Forty-five patients were male. Several tumor locations were included in the study: tongue (n=58), floor of the mouth (n=24), palate (n=2), and the lip (n=1). The 7th and 8th editions of the AJCC staging system were used. Tumor thickness was measured with ultrasound in 46 patients and with MR imaging in 76 patients. For ultrasound, the probe (13 to 7 MHz transducer) was placed directly on the lesion. MR imaging was performed and tumor dimensions were measured in 3D (used sequences: T1W, TSE, TRA, TR, TE 538/10ms, flip angle 90, matrix 288/248, slice thickness of 4mm, STIR TSE COR, TR/TE 2500/60ms, matrix 216/170, T1 3D Thrive fat-saturation, intravenous injection of 15cc gadoterate meglumine, TR/TE 9.86/4.59ms, flip angle 10, matrix 200/179, slice thickness 1mm). Radiologists reported the MRI outcome and suggested a T-stage. The pathologist reported the tumor dimension in the pathological report. Further pathological procedures were not described and it was unclear how

specimens were fixed. The time period between the preoperative assessments and the histopathological assessment was unclear.

Choi (2017) assessed the agreement between clinical examination and histopathology for the tumor thickness (categorized in T-stages) in n=252 patients with biopsy proven squamous cell carcinomas of the oral cavity. Patients had a median age of 55 years (range: 47 to 65) and had tumors on the tongue (n=195), floor of mouth (n=34), or on the buccal mucosa (n=23). Various T-stages were included in the sample: pT1 (n=109), pT2 (n=80), pT3 (n=25), pT4a (n=37), and pT4b (n=1). The 7th edition of the AJCC staging system was used. Clinical examination consisted of a physician performing preoperative endoscopic assessment, palpation and imaging with either CT or MRI. Surgical specimens of the primary tumor were assessed microscopically, however it remained unclear how specimens were fixed. Agreement between the T-stages as assessed with the preoperative clinical examination and postoperative histopathology was quantified with a Cohen's Kappa. The time period between both assessments was unclear.

Klein Nulent (2018) performed a systematic review with a search up to the 6th of July 2016 in PubMed (Medline), EMBASE, and the Cochrane databases for studies comparing intraoral ultrasound tumor thickness measurements with postoperative pathological assessment. Included studies had to contain patients with oral squamous cell carcinoma and the ultrasound measurements had to be performed preoperatively or intraoperatively. Included patients had tongue tumors, buccal mucosa tumors, tumors on the floor of the mouth, lip tumours, or alveolar mucosa tumors. Data was extracted according to the 7th edition of the AJCC staging system. Ten out of the twelve included studies (n=240) were used for the assessment of the agreement between intraoral ultrasound and postoperative histopathology in a Bland-Altman plot. For one of the included studies the authors estimated the individual patient data from a figure. The QUADAS-2 tool was used to assess the risk of bias of the included studies. Time between preoperative measurement and postoperative histology was not assessed. The tissue fixation method for histopathological analyses were not reported for the individual studies included in the systematic review.

Nair (2018) recruited patients with biopsy proven T1N0 (n=18) or T2N0 (n=6) primary squamous cell carcinomas of the tongue to assess the agreement between ultrasound and histopathology for tumor thickness. A total of 25 patient were recruited with a median age of 55 years (range: 22 to 76). Sixteen of the recruited patients were male. The 6th edition of the AJCC staging system was used. Preoperative ultrasound assessments using a 17 or 9 MHz conventional linear probe were performed with patients extending their tongue. The probe was placed directly upon the lesion. The surgical tumor specimens were placed in saline and immediately send to the pathology department for assessment (specimens were not fixed in formalin), where the specimens were cut into 2 to 3 millimeter thick transverse slices. The time period between the preoperative ultrasound assessment and the histopathological assessment was unclear.

Shintani (2001) assessed the agreement between CT or MRI and histopathology for the measurement of tumor thickness in 38 patients with oral cancer. Furthermore, ultrasound (7.5-Mhz intracavitarytransducers) was assessed and this data was included in the systematic review of Klein Nulent (2018). The patients had a mean age of 58.2 years (SD: not reported, range: 36 to 91) and had tumors on the tongue (n=26), buccal mucosa (n=8), and floor of mouth (n=4). The 5th edition of the UICC TNM staging system was used. Tumor thickness was measured with a contrast-enhanced 5-mm axial CT in 38 participants and with 4-mm axial and coronal T1/T2-weighted MR imaging in 26 patients. Histological sections

(presumably from formalin fixed tissue) were assessed with a micrometer. The authors state that tumors smaller than 5 millimeters were difficult to differentiate with CT or MRI. Tumors were not detected by CT in 19 patients and by MRI in 11 patients. These patients were therefore not included in the analyses. Furthermore, the time period between the preoperative assessment and the histopathologic assessment was unclear.

Results

Depth of invasion

Results concerning the instrument agreement for depth of invasion are summarized in Table 1.1.

Categorical (T-stage)

Clinical examination

Goel (2016) reported the agreement between a clinical examination and histopathology in T-stages in 63 patients. It was unclear what the procedures for clinical examination were. Clinical assessment for T-stage showed an agreement of Cohen's Kappa = 0.47 (95%CI: not reported) with histopathology.

CT

No studies were included that reported the T-stage agreement between CT and histopathology while measuring depth of invasion.

PET-CT

No studies were included that reported the T-stage agreement between PET-CT and histopathology while measuring depth of invasion.

MRI

Goel (2016) reported the agreement between MRI and histopathology in T-stages in 63 patients. A Cohen's Kappa of 0.69 (95%CI: not reported) was found.

Verma (2019) reported the T-stage classifications of MRI and histopathology in a three-by-thee table (T1-3) from 50 included patients. A Cohen's Kappa was not reported but could be calculated from the table. Here, a Kappa of 0.65 was calculated (95%CI: not calculated).

Vidiri (2019) reported the T-stage agreement of an experienced and an inexperienced radiologist interpreting MR imaging with the histopathology results. The experienced radiologist showed a Cohen's Kappa of 0.74 (95%CI: 0.56 to 0.92) for the agreement of T-stage between MRI and histopathology. For the inexperienced radiologist the Cohen's Kappa was 0.60 (95%CI: 0.40 to 0.80).

Ultrasound

No studies were included that reported the T-stage agreement between ultrasound and histopathology while measuring depth of invasion.

Categorical (at a threshold)

Clinical examination

Alsaffar (2016) categorized the depth of invasion, which resulted in two categories: < 5 millimeters depth of invasion and ≥ 5 millimeters depth of invasion. The treating surgeon performed a palpation to assess the depth of invasion. The agreement between the treating surgeon's preoperative palpation and the postoperative histopathology was quantified with a Cohen's Kappa (n=53). A Kappa of 0.61 (95%CI: 0.36 to 0.87) was reported.

CT
No studies were included that reported the agreement at a specified threshold between CT and histopathology while measuring depth of invasion.

PET-CT
No studies were included that reported the agreement at a specified threshold between PET-CT and histopathology while measuring depth of invasion.

MRI
Alsaffar (2016) assessed the agreement between MRI and histopathology for measuring depth of invasion and used two categories: < 5 millimeters and ≥ 5 millimeters. A Cohen's Kappa of 0.80 (95%CI: 0.59 to 1.00) was reported (n=43).

Ultrasound
Iida (2018) assessed the agreement between ultrasound and histopathology for the depth of invasion in 53 participants. Depth of invasion was categorized in: < 5 millimeters and ≥ 5 millimeters. A Cohen's Kappa of 0.65 (95%CI: 0.43 to 0.87) was reported.

Continuous (in millimeters)
Clinical examination
No studies were included that reported the agreement on a continuous scale between clinical examination (palpation) and histopathology while measuring depth of invasion.

CT
No studies were included that reported the agreement on a continuous scale between CT and histopathology while measuring depth of invasion.

PET-CT
No studies were included that reported the agreement on a continuous scale between PET-CT and histopathology while measuring depth of invasion.

MRI
Mao (2019) constructed Bland-Altman plots for the agreement between MRI and histopathology measuring the depth of invasion. Several plots were constructed for different tumor stages and types. Overall, MRI showed a mean overestimation of 2.32 millimeters when compared to the histopathologic results (n=150). In 95% of the measurements, MRI measured between an underestimation of 0.97 millimeters (-0.97 millimeters) and an overestimation of 5.61 millimeters. Furthermore, agreement per tumor stage was assessed: T1 (mean difference: 1.46 millimeters, 95% limits of agreement: -0.67 to 3.63 millimeters, n=43), T2 (mean difference: 2.08 millimeters, 95% limits of agreement: -0.45 to 4.62 millimeters, n=71), and T3 (mean difference: 3.79 millimeters, 95% limits of agreement: -0.13 to 7.7 millimeters, n=36). Finally, agreement per tumor type was assessed: ulcer type (mean difference: 3.72 millimeters, 95% limits of agreement: -0.16 to 7.6 millimeters, n=41), invasive type (mean difference: 1.83 millimeters, 95% limits of agreement: -0.59 to 4.25 millimeters, n=91), and exogenous type (mean difference: 1.53 millimeters, 95% limits of agreement: 0.01 to 3.06 millimeters, n15).

Vidiri (2019) constructed Bland-Altman plots for the agreement between an experienced or inexperienced radiologist using MRI and histopathology for the depth of invasion in 43 patients. The MRI measurements by an experienced radiologist had a mean underestimation of 0.3 millimeters (-0.3 millimeters), where 95% of the measurements lay between an underestimation of 5.5 millimeters (-5.5 millimeters) and an overestimation of 4.9 millimeters. For the inexperienced radiologist the MRI measurements had a mean underestimation of 0.4 millimeters (-0.4 millimeters), while 95% of the MRI measurements lay between an underestimation of 6.6 millimeters (-6.6 millimeters) and 5.8 millimeters overestimation.

Ultrasound
No studies were included that reported the agreement on a continuous scale between ultrasound and histopathology while measuring depth of invasion.

Table 1.1 Study results for depth of invasion per measurement level per instrument

Variable	Measurement level	Measurement instrument	Threshold	Author	Result	Risk of Bias (COSMIN, unless stated otherwise)
Depth of invasion	**Categorical (T-stage)**	Clinical examination (unclear procedures)	T-stage	Goel 2016	Kappa for the agreement of T-stage, n=61 (clinical examination versus pathological data): K=0.47 (95%CI: not reported)	Doubtful
		CT	No studies were included that reported the T-stage agreement between CT and histopathology while measuring depth of invasion.			
		PET-CT	No studies were included that reported the T-stage agreement between PET-CT and histopathology while measuring depth of invasion.			
		MRI	T-stage	Goel 2016	Kappa for the agreement of T-stage, n=61 (MRI versus pathological data): K=0.69 (95%CI: not reported)	Doubtful
			T-stage	Verma 2019	Kappa for the agreement of tumour thickness (T-stage) as measured by MRI and histopathology was not reported. However, the 3x3 table was reported from which a Kappa could be calculated in n= 50: Kappa = 0.65	Doubtful
			T-stage	Vidiri 2019	Kappa for the agreement of T-stage, n=43 (MRI experienced radiologist versus pathological data): K=0.74 (95%CI: 0.56-0.92)	Adequate

					Kappa for the agreement of T-stage, n=43 (MRI inexperienced radiologist versus pathological data): K=0.60 (95%CI: 0.40-0.80)	
		Ultrasound	T-stage	No studies were included that reported the T-stage agreement between ultrasound and histopathology while measuring depth of invasion.		
	Categorical (at a threshold)	Clinical examination (palpation)	5 mm	Alsaffar 2016	Kappa at a threshold of 5 millimetres, n=53: K=0.61 (95%CI: 0.36-0.87)	Doubtful
		CT	No studies were included that reported the agreement at a specified threshold between CT and histopathology while measuring depth of invasion.			
		PET-CT	No studies were included that reported the agreement at a specified threshold between PET-CT and histopathology while measuring depth of invasion.			
		MRI	5 mm	Alsaffar 2016	Kappa at a threshold of 5 millimetres, n=53: K=0.80 (95%CI: 0.59-1.00)	Doubtful
		Ultrasound	5 mm	Iida 2018	Kappa at a threshold of 5 millimetres, n=59: K=0.651 (95%CI: 0.43-0.87)	Doubtful
	Continuous	Clinical examination	No studies were included that reported the agreement on a continuous scale between clinical examination (palpation) and histopathology while measuring depth of invasion.			
		CT	No studies were included that reported the agreement on a continuous scale between CT and histopathology while measuring depth of invasion.			
		PET-CT	No studies were included that reported the agreement on a continuous scale between PET-CT and histopathology while measuring depth of invasion.			
		MRI	NA	Mao 2019	Bland-Altman plot overall n=150 (MRI-histopathology), mm: Mean difference: 2.32. 95% upper limit: 5.61 95% lower limit: -0.97 Bland-Altman plot tumour T1-stage n=43 (MRI-histopathology), mm: Mean difference: 1.48. 95% upper limit: 3.63 95% lower limit: -0.67 Bland-Altman plot tumour T2-stage n=71 (MRI-histopathology), mm: Mean difference: 2.08. 95% upper limit: 4.62 95% lower limit: -0.45	Adequate

					Bland-Altman plot tumour T3-stage n=36 (MRI-histopathology), mm: Mean difference: 3.79. 95% upper limit: 7.70 95% lower limit: -0.13 Bland-Altman plot ulcer type tumour n=41 (MRI-histopathology), mm: Mean difference: 3.72. 95% upper limit: 7.60 95% lower limit: -0.16 Bland-Altman plot invasive type tumour n=91 (MRI-histopathology), mm: Mean difference: 1.83. 95% upper limit: 4.25 95% lower limit: -0.59 Bland-Altman plot exogenous type tumour n=15 (MRI-histopathology), mm: Mean difference: 1.53. 95% upper limit: 3.06 95% lower limit: 0.01	
			NA	Vidiri 2019	Bland-Altman plot n=43 (MRI experienced radiologist-histopathology), mm: Mean difference: -0.3 95% upper limit: 4.9 95% lower limit: -5.5 Bland-Altman plot n=43 (MRI inexperienced radiologist-histopathology), mm: Mean difference: -0.4 95% upper limit: 5.8 95% lower limit: -6.6	Adequate
		Ultrasound	No studies were included that reported the agreement on a continuous scale between ultrasound and histopathology while measuring depth of invasion.			
NA: Not Applicable						

Tumor thickness

Results concerning the instrument agreement for tumor thickness are summarized in Table 1.2.

Categorical (T-stage)

Clinical examination

Choi (2017) reported a Cohen's Kappa = 0.81 (95%CI not reported) for the agreement of T-stages between a preoperative clinical assessment and a histopathologic assessment in 252 participants. The clinical examination consisted of an endoscopic assessment, a palpation, and either CT or MR imaging.

CT

No studies were included that reported the T-stage agreement between CT and histopathology while measuring tumor thickness.

PET-CT

No studies were included that reported the T-stage agreement between PET-CT and histopathology while measuring tumor thickness.

MRI

No studies were included that reported the T-stage agreement between MRI and histopathology while measuring tumor thickness.

Ultrasound

No studies were included that reported the T-stage agreement between ultrasound and histopathology while measuring tumor thickness.

Categorical (at a threshold)

Clinical examination

No studies were included that reported the agreement at a specified threshold between clinical examination and histopathology while measuring tumor thickness.

CT

No studies were included that reported the agreement at a specified threshold between CT and histopathology while measuring tumor thickness.

PET-CT

No studies were included that reported the agreement at a specified threshold between PET-CT and histopathology while measuring tumor thickness.

MRI

No studies were included that reported the agreement at a specified threshold between MRI and histopathology while measuring tumor thickness.

Ultrasound

No studies were included that reported the agreement at a specified threshold between ultrasound and histopathology while measuring tumor thickness.

Continuous (in millimeters)

Clinical examination

No studies were included that reported the agreement on a continuous scale between clinical examination (palpation) and histopathology while measuring tumor thickness.

CT
Shintani (2001) did not report agreement parameters. However, the agreement could be calculated from the reported individual patient data (n=19). CT had a mean overestimation of 5.93 millimeters. When the 95% limits of agreement were calculated, 95% of the CT measurements lay between an underestimation of 5.66 millimeters (-5.66 millimeters) and an overestimation of 17.53 millimeters compared to histopathology.

PET-CT
No studies were included that reported the agreement on a continuous scale between PET-CT and histopathology while measuring tumor thickness.

MRI
Brouwer de Koning (2019) constructed a Bland-Altman plot where the mean overestimation of MRI was 1.3 millimeters in 83 patients. Ninety-five percent of the MRI measurements fell between an underestimation of 6.1 millimeters (-6.1 mm) and an overestimation of 8.6 millimeters compared to histopathology.

Shintani (2001) did not report agreement parameters. Nonetheless, the agreement between MRI and histopathology could be calculated (n=13). The mean difference was an overestimation of 8.55 millimeters by MRI. When the 95% limits of agreement were calculated, 95% of the MRI measurements lay between an underestimation of 5.94 millimeters (-5.94 millimeters) and an overestimation of 23.05 millimeters compared to histopathology.

Ultrasound
Brouwer de Koning (2019) reported a mean overestimation of 0.05 millimeters by ultrasound when compared to histopathological results in 83 patients. The ultrasound measurements were in 95% of the cases between an underestimation of 5.3 millimeters (-5.3 millimeters) and an overestimation of 5.4 millimeters when compared to histopathologic results.

Klein Nulent (2018) performed a systematic review and used individual patient data from 240 patients to construct a Bland-Altman plot. Ultrasound had a mean overestimation of 0.5 millimeters compared to histopathology. In 95% of the measurements the ultrasound resulted in measurements between -5.5 millimeters (5.5 millimeters underestimation) and 6.5 millimeters (6.5 millimeters overestimation) when compared to histopathology results.

Nair (2018) recruited 24 patients for the agreement between ultrasound and histopathology measuring tumor thickness. A Bland-Altman plot showed a mean difference between ultrasound an histoplathology where ultrasound underestimated the tumor thickness by 0.15 millimeters (-0.15 millimeters). The limits of agreement were not reported but could be approximated from the reported figure. Here, 95% of the ultrasound measurements were between 4.6 millimeters underestimation (-4.6 millimeters) and 4.99 millimeters overestimation compared to histopathologic results.

Table 1.2 Study results for tumor thickness per measurement level per instrument

Variable	Measurement level	Measurement instrument	Threshold	Author	Result	COSMIN Risk of Bias
Tumor thickness	**Categorical (T-stage)**	Clinical examination (endoscopic + palpation + CT or MRI)	T-stage	Choi 2017	Kappa for the agreement of T-stage n=252 (clinical examination	Doubtful

					versus pathological data): K=0.81 (95%CI: not reported)	
		CT	No studies were included that reported the T-stage agreement between CT and histopathology while measuring tumor thickness.			
		PET-CT	No studies were included that reported the T-stage agreement between PET-CT and histopathology while measuring tumor thickness.			
		MRI	No studies were included that reported the T-stage agreement between MRI and histopathology while measuring tumor thickness.			
		Ultrasound	No studies were included that reported the T-stage agreement between ultrasound and histopathology while measuring tumor thickness.			
	Categorical (at a threshold)	Clinical examination	No studies were included that reported the agreement at a specified threshold between clinical examination and histopathology while measuring tumor thickness.			
		CT	No studies were included that reported the agreement at a specified threshold between CT and histopathology while measuring tumor thickness.			
		PET-CT	No studies were included that reported the agreement at a specified threshold between PET-CT and histopathology while measuring tumor thickness.			
		MRI	No studies were included that reported the agreement at a specified threshold between MRI and histopathology while measuring tumor thickness.			
		Ultrasound	No studies were included that reported the agreement at a specified threshold between ultrasound and histopathology while measuring tumor thickness.			
	Continuous	Clinical examination	No studies were included that reported the agreement on a continuous scale between clinical examination (palpation) and histopathology while measuring tumor thickness.			
		CT	NA	Shintani 2001	Bland-Altman parameters calculated from presented data n=19 (CT-histopathology), mm: Mean difference: 5.93. 95% upper limit: 17.53 95% lower limit: -5.66	Doubtful
		PET-CT	NA	No studies were included that reported the agreement on a continuous scale between PET-CT and histopathology while measuring tumor thickness.		
		MRI	NA	Brouwer de Koning 2019	Bland-Altman plot n=83 (MRI-histopathology), mm: Mean difference: 1.3. 95% upper limit: 8.6 95% lower limit: -6.1	Doubtful

			NA	Shintani 2001	Bland-Altman parameters calculated from presented data n=13 (MRI-histopathology), mm: Mean difference: 8.55. 95% upper limit: 23.05 95% lower limit: -5.94	Doubtful
		Ultrasound	NA	Brouwer de Koning 2019	Bland-Altman plot n=83 (US-histopathology), mm: Mean difference: 0.05. 95% upper limit: 5.4 95% lower limit: -5.3	Doubtful
			NA	Klein Nulent 2018	Bland-Altman plot n=240 (ultrasound-histopathology), mm: Mean difference: 0.5. 95% upper limit: 6.5 95% lower limit: -5.5	Klein Nulent 2018 assessed the risk of bias with the QUADAS-2 tool. For flow and timing: 4 low risk / 1 high risk / 7 unclear
			NA	Nair 2018	Bland-Altman plot overall n=150 (US-histopathology), mm: Mean difference: -0.15 95% upper limit: 4.99 95% lower limit: -4.6 Limits of agreement were approximated from the provided Bland-Altman plot:	Doubtful
NA: Not applicable						

<u>Level of evidence of the literature</u>

Depth of invasion

<u>Categorical (T-stage)</u>

Clinical examination

The level of evidence regarding clinical examination for the outcome measure 'categorical agreement (T-stage)' was downgraded by 3 levels because of study limitations (2 level for risk of bias: there is only one study of doubtful quality), and the number of included patients (1 level for imprecision: sample size was less than 100, but more than 50); publication bias was not assessed.

CT
GRADE could not be applied because none of the included studies reported data about the categorical agreement on T-stage between CT and histopathology when measuring depth of invasion.

PET-CT
GRADE could not be applied because none of the included studies reported data about the categorical agreement on T-stage between PET-CT and histopathology when measuring depth of invasion.

MRI
The level of evidence regarding MRI for the outcome measure 'categorical agreement (T-stage)' was downgraded by 1 level because of study limitations (1 level for risk of bias: multiple studies of doubtful quality and one study of adequate quality); publication bias was not assessed.

Ultrasound
GRADE could not be applied because none of the included studies reported data about the categorical agreement on T-stage between ultrasound and histopathology when measuring depth of invasion.

Categorical (at a threshold)
Clinical examination
The level of evidence regarding clinical examination (palpation) for the outcome measure 'categorical agreement (at a threshold)' was downgraded by 3 levels because of study limitations (2 levels for risk of bias: there is only one study of doubtful quality) and the number of included patients (1 level for imprecision: sample size was less than 100, but more than 50); publication bias was not assessed.

CT
GRADE could not be applied because none of the included studies reported data about the categorical agreement for depth of invasion at a threshold between CT and histopathology.

PET-CT
GRADE could not be applied because none of the included studies reported data about the categorical agreement for depth of invasion at a threshold between PET-CT and histopathology.

MRI
The level of evidence regarding MRI for the outcome measure 'categorical agreement (at a threshold)' was downgraded by 3 levels because of study limitations (2 levels for risk of bias: there is only one study of doubtful quality) and the number of included patients (1 level for imprecision: sample size was less than 100, but more than 50); publication bias was not assessed.

Ultrasound
The level of evidence regarding ultrasound for the outcome measure 'categorical agreement (at a threshold)' was downgraded by 3 levels because of study limitations (2 levels for risk of bias: there is only one study of doubtful quality) and the number of included patients (1 level for imprecision: sample size was less than 100, but more than 50); publication bias was not assessed.

Continuous (in millimeters)
Clinical examination
GRADE could not be applied because none of the included studies reported data about the agreement on a continuous scale for depth of invasion between a clinical examination and histopathology.

CT
GRADE could not be applied because none of the included studies reported data about the agreement on a continuous scale for depth of invasion between CT and histopathology.

PET-CT
GRADE could not be applied because none of the included studies reported data about the agreement on a continuous scale for depth of invasion between PET-CT and histopathology.

MRI
The level of evidence regarding MRI for the outcome measure 'continuous agreement (in millimeters)' was downgraded by 1 level because of conflicting results (1 level for inconsistency: Mao (2019) reports a mean overestimation of 2.32 millimeters, while Vidiri (2019) reports a mean underestimation of 0.3 millimeters. Furthermore, Vidiri (2019) reports wider 95% lower limits of agreement when compared to Mao (2019): an underestimation of 5.5 millimeters (Vidiri, 2019) versus an underestimation of 0.97 millimeters (Mao, 2019)); publication bias was not assessed.

Ultrasound
GRADE could not be applied because none of the included studies reported data about the agreement on a continuous scale for depth of invasion between ultrasound and histopathology.

Tumor thickness
Categorical (T-stage)
Clinical examination
The level of evidence regarding a clinical examination (consisting of an endoscopic examination, palpation, and either CT or MR imaging) for the outcome measure 'categorical agreement (T-stage)' was downgraded by two levels because of study limitations (2 levels for risk of bias: there is only one study of doubtful quality); publication bias was not assessed.

CT
GRADE was not applied because none of the included studies reported data about the categorical agreement on T-stage between CT and histopathology when measuring tumor thickness.

PET-CT
GRADE was not applied because none of the included studies reported data about the categorical agreement on T-stage between PET-CT and histopathology when measuring tumor thickness.

MRI
GRADE was not applied because none of the included studies reported data about the categorical agreement on T-stage between MRI and histopathology when measuring tumor thickness.

Ultrasound
GRADE was not applied because none of the included studies reported data about the categorical agreement on T-stage between ultrasound and histopathology when measuring tumor thickness.

Categorical (at a threshold)
Clinical examination
GRADE was not applied because none of the included studies reported data about the categorical agreement for tumor thickness at a threshold between a clinical examination and histopathology.

CT
GRADE was not applied because none of the included studies reported data about the categorical agreement for tumor thickness at a threshold between CT and histopathology.

PET-CT
GRADE was not applied because none of the included studies reported data about the categorical agreement for tumor thickness at a threshold between PET-CT and histopathology.

MRI
GRADE was not applied because none of the included studies reported data about the categorical agreement for tumor thickness at a threshold between MRI and histopathology.

Ultrasound
GRADE was not applied because none of the included studies reported data about the categorical agreement for tumor thickness at a threshold between ultrasound and histopathology.

Continuous (in millimeters)
Clinical examination
GRADE was not applied because none of the included studies reported data about the agreement on a continuous scale for tumor thickness between a clinical examination and histopathology.

CT
The level of evidence regarding CT for the outcome measure 'agreement on a continuous measurement level (in millimeters)' was downgraded by 4 levels because of study limitations (2 levels for risk of bias: there is only one study of doubtful quality) and the number of included patients (2 levels for imprecision: the sample size was less than 50); publication bias was not assessed.

PET-CT
GRADE was not applied because none of the included studies reported data about the agreement on a continuous scale for tumor thickness between PET-CT and histopathology.

MRI
The level of evidence regarding MRI for the outcome measure 'agreement on a continuous measurement level (in millimeters)' was downgraded by 2 levels because of study limitations (1 level for risk of bias: there were multiple studies of doubtful quality) and the number of

included patients (1 level for imprecision: the sample size was less than 100, but more than 50); publication bias was not assessed.

Ultrasound
The level of evidence regarding ultrasound for the outcome measure 'agreement on a continuous measurement level (in millimeters)' was downgraded by 1 level because of study limitations (1 level for risk of bias: there were multiple studies of doubtful quality. Klein Nulent 2018 assessed the risk of bias with the QUADAS-2 and scored 4 studies with low risk / 1 study with high risk / 7 studies with unclear risk on the 'flow and timing' item); publication bias was not assessed.

Conclusions
Depth of invasion
The agreement estimates of modalities measuring depth of invasion and their certainty (following GRADE) are summarized in Table 1.3.

Table 1.3 Summarized results for the agreement and GRADE certainty of clinical examination, CT, PET-CT, MRI, or intraoral ultrasound measuring depth of invasion

Modality	Agreement on a categorical level per T-Stage (GRADE certainty)	Agreement on a categorical level using a threshold (GRADE certainty)	Agreement on a continuous level (GRADE certainty)
Clinical examination	K = 0.47 (unclear procedures) (VERY LOW)	K = 0.61 (95%CI: 0.36-0.87) at a 5-millimeter threshold (palpation) (VERY LOW)	NA
	References: Goel 2016	*References: Alsaffar 2016*	*NA*
CT	NA	NA	NA
PET-CT	NA	NA	NA
MRI	Range: K = 0.60-0.74 (MODERATE)	K = 0.80 (95%CI: 0.59-1.00) at a 5-millimeter threshold (VERY LOW)	Range upper 95% LoA: 4.9-5.8* Range mean difference: -0.4–2.32* Range lower 95% LoA: -0.97– -6.6* (MODERATE)
	References: Goel 2016; Verma 2019; Vidiri 2019	*References: Alsaffar 2016*	*References: Mao 2019; Vidiri 2019*
Ultrasound	NA	K = 0.65 (95%CI: 0.43-0.87) at a 5-millimeter threshold (VERY LOW)	NA
	NA	*References: Iida 2018*	*NA*
***Sub-analyses in Mao 2019 were not included in the range** **CI: Confidence Interval** **LoA: Limit of Agreement** **NA: Not Available**			

Tumor thickness
The agreement estimates of modalities measuring tumor thickness and their certainty (following GRADE) are summarized in Table1.4.

Table1.4 Summarized results for the agreement and GRADE certainty of clinical examination, CT, PET-CT, MRI, or intraoral ultrasound measuring tumor thickness

Modality	Agreement on a categorical level per T-Stage (GRADE certainty)	Agreement on a categorical level using a threshold (GRADE certainty)	Agreement on a continuous level (GRADE certainty)
Clinical examination	K = 0.81 (endoscopic examination, palpation, and either CT or MR imaging) (LOW)	NA	NA
	References: *Choi 2017*	*NA*	*NA*
CT	NA	NA	Upper 95% LoA: 17.53 Mean difference: 5.93 Lower 95% LoA: -5.66 (VERY LOW)
	NA	NA	*References:* Shintani 2001
PET-CT	NA	NA	NA
MRI		NA	Range upper 95% LoA: 8.6-23.05 Range mean difference: 1.3-8.55 Range lower 95% LoA: -5.94– -6.1 (LOW)
	NA	*NA*	*References:* *Brouwer de Koning 2019; Shintani 2001*
Ultrasound	NA	NA	Range upper 95% LoA: 4.99-6.5 Range mean difference: -0.15-0.5 Range lower 95% LoA: -4.6– -5.5 (MODERATE)
	NA	*NA*	*References:* *Brouwer de Koning 2019; Klein Nulent 2018; Nair 2018*
CI: Confidence Interval **LoA: Limit of Agreement** **NA: Not Available**			

Overwegingen - van bewijs naar aanbeveling

De gevonden resultaten met betrekking tot invasiediepte zijn samengevat in Tabel 1.1, welke is te vinden in de samenvatting van de literatuur (onder het tabblad 'onderbouwing'). Voor de overeenkomst op T-stadium met behulp van invasiedieptemetingen ten opzichte van histopathologie werd er geen data gevonden voor het gebruik van CT en PET-CT. Voor de overeenkomst, gecategoriseerd door een afkapwaarde, in invasiedieptemetingen werd er geen data gevonden die CT of PET-CT vergeleek met histopathologie. Voor de

overeenstemming van invasiedieptemetingen op een continue schaal, in millimeters, werden ten slotte geen data gevonden voor CT, PET-CT en ultrageluid.

De bewijskracht werd met een aangepaste versie van GRADE beoordeeld (Mokkink, 2018). Voor de overeenstemming op T-stadium waarbij invasiedieptemetingen werden gebruikt was het vertrouwen, volgens de GRADE, in een klinische beoordeling (met onduidelijke procedures) laag. Goel (2016) rapporteerde een overeenstemming van K = 0,47 tussen een klinische beoordeling (met onduidelijke procedures) en histopathologie. Het vertrouwen in de gerapporteerde uitkomsten van MRI voor de overeenstemming op T-stadium was redelijk. Hier rapporteerden drie studies (Goel, 2016; Verma, 2019; Vidiri, 2019) hun uitkomsten. Verma (2019) en Vidiri (2019) gebruikten in de studies de 8^{e} editie van de TNM-classificering (Verma (2019) gebruikte óók de 7^{e} editie). Er werden Kappa-waarden van 0.69 (95%BHI: niet gerapporteerd), 0.65 (95%BHI: niet berekend), 0.74 (ervaren beoordelaar, 95%BHI: 0.56 tot 0.92) en 0.60 (onervaren beoordelaar, 95%BHI: 0.40 tot 0.80) door de studie-auteurs gerapporteerd (Goel, 2016; Verma, 2019; Vidiri, 2019). De gevonden resultaten liggen rond de vooraf gedefinieerde grens van besluitvorming (dat wil zeggen dat K ≥ 0.70 als voldoende overeenstemming werd gezien).

Twee studies werden geïncludeerd voor invasiedieptemetingen die gecategoriseerd werden aan de hand van een afkappunt (Alsaffar, 2016; Iida, 2018). De enige afkapwaarde die werd gebruikt was 5 millimeter waardoor er twee categorieën ontstonden (dat wil zeggen een invasiediepte van < 5 millimeter of ≥ 5 millimeter). Het vertrouwen in de gerapporteerde uitkomsten waren volgens GRADE zeer laag voor een klinische beoordeling, MRI en ultrageluid. De zeer lage GRADE-beoordeling ontstond vooral vanwege het risico op vertekening van de resultaten en de beperkte omvang van de steekproeven. In beide studies was de periode tussen het preoperatieve assessment en het histopathologische assessment onduidelijk.

Ten slotte werd er voor de overeenstemming van invasiedieptemetingen op een continue schaal (in millimeters) tussen modaliteiten en histopathologie alleen data gevonden voor MRI (Mao, 2019; Vidiri, 2019). Beide studies rapporteerden de tijdsperiode tussen het preoperatieve assessment en het histopathologisch assessment deels. Mao (2019) beschreef dat de preoperatieve meting met MRI binnen 1 week vóór resectie werd uitgevoerd, terwijl dit 3 tot 4 weken was voor Vidiri (2019). De werkgroep had a priori vastgesteld dat een meting tot maximaal 4 weken vóór de chirurgische resectie als adequaat werd gezien. Voor het meten van de invasiediepte op continue schaal (in millimeters) met MRI is er, volgens GRADE, een redelijk vertrouwen in de gevonden resultaten voor MRI uit de twee studies. In de gerapporteerde data werd gezien dat 95% van de MRI metingen (n=150) tussen een onderschatting van 0,97 millimeter en een overschatting van 5,61 millimeter lag in één van de studies (Mao, 2019). In de andere studie (n=53) lagen 95% van de metingen tussen 5,5 millimeter onderschatting en 4,9 millimeter overschatting door een ervaren radioloog en tussen 6,6 millimeter onderschatting en 5,8 millimeter overschatting door een onervaren radioloog (Vidiri, 2019).

De gevonden resultaten met betrekking tot tumordikte zijn samengevat in Tabel 1.2 en is te vinden in de samenvatting van de literatuur (onder het tabblad ‘onderbouwing’). Voor de overeenkomst op T-stadium met behulp van tumor diktemetingen werd geen data gevonden voor het gebruik van CT, PET-CT, MRI en ultrageluid. Voor de overeenkomst, gecategoriseerd door een afkapwaarde, in tumor diktemetingen werd er voor geen enkele modaliteit van interesse data gevonden (dat wil zeggen klinisch onderzoek, CT, PET-CT, MRI

en ultrageluid). Voor de overeenstemming van invasiedieptemetingen op een continue schaal, in millimeters, werden ten slotte enkel voor klinisch onderzoek geen data gevonden.

Voor de overeenstemming op T-stadium met behulp van tumor diktemetingen werd er één studie geïncludeerd (Choi, 2017). Deze studie beschreef een work-up (bestaande uit een endoscopisch beoordeling, palpatie en beeldvorming door CT of MRI) ten opzichte van een histopathologische beoordeling. Er werd een Kappa-waarde van 0.80 gerapporteerd, maar het 95%BHI werd niet vermeld en het vertrouwen in deze uitkomst was laag (volgens GRADE).

Er werden geen studies geïncludeerd die de overeenstemming voor tumordikte categoriseerden met behulp van specifieke afkapwaarden. Hierdoor is er voor geen enkele modaliteit van interesse data voor deze specifieke situatie beschikbaar.

Wanneer de overeenstemming op een continue schaal (in millimeters) werd onderzocht was het vertrouwen in de gerapporteerde resultaten, volgens GRADE, in CT zeer laag, in MRI laag en in ultrageluid redelijk. Het vertrouwen in de gerapporteerde uitkomsten van MRI werd verlaagd vanwege de beperkte steekproefomvang en het risico op vertekening van de resultaten. De tijdsperiode tussen het preoperatieve assessment en het histopathologische assessment was onduidelijk in beide studies. Bij het gebruik van MRI lagen 95% van de metingen tussen een onderschatting van 5,3 millimeter en een overschatting van 5,4 millimeter bij 83 participanten (Brouwer de Koning, 2019) en tussen een onderschatting van 4,6 en een overschatting van 4,99 millimeter bij 150 patiënten (Nair, 2018) ten opzichte van histopathologie. Er was een redelijk vertrouwen in de gerapporteerde uitkomsten van ultrageluid. Klein Nulent (2018) combineerde data van 10 studies waardoor de gepresenteerde Bland-Altman plots data van 240 patiënten bevatte. Er werd gerapporteerd dat 95% van de metingen met ultrageluid tussen een onderschatting van 5,5 millimeter en een overschatting van 6,5 millimeter lagen ten opzichte van histopathologie (Klein Nulent, 2018).

Een onderschatting van de invasiediepte of tumordikte geeft een verhoogd risico op inadequate snijranden, terwijl een overschatting een verhoogd risico op te ruime resectieranden geeft. Als de resectieranden inadequaat zijn is er (vaak in combinatie met andere negatieve histopathologische bevindingen) een indicatie voor adjuvante therapie in de vorm van een heroperatie of radiotherapie met of zonder chemotherapie. Door inadequate resectieranden kan de overleving verminderd zijn. Als gevolg van adjuvante behandeling of te ruime resectieranden kunnen mondfuncties zijn aangedaan en kwaliteit van leven verminderd zijn. In het algemeen wordt onderschatting ernstiger gevonden dan overschatting.

De werkgroep werd na de zoekdatum voor literatuur op de hoogte gesteld van een aantal relevante publicaties. Deze publicaties werden daarom niet in de literatuuranalyse opgenomen, maar zullen kort besproken worden ter overweging. Deze korte bespreking geeft wellicht geen compleet literatuuroverzicht van de periode na de systematische zoekopdracht in deze richtlijnmodule en bevat geen GRADE-beoordelingen.

Noorlag (2020) onderzocht met retrospectieve data de tumordiepte gemeten met MRI of intra-orale echografie ten opzichte van een postoperatief histopathologisch assessment. De auteurs rapporteerden een Pearson's correlatiecoëfficiënt van 0.792 ($p<0,001$) voor MRI ten opzichte van histopathologie. Er werd ook bekeken welke beeldvorming een groter verschil ten opzichte van histopathologie had bij tumoren met een kleine invasiediepte (≤ 1

centimeter) en een grote invasiediepte (> 1 centimeter). De auteurs concludeerden dat intra-orale echografie voor tumoren met een kleine invasiediepte accurater zou zijn dan MRI, maar dat echografie bij dikkere tumoren de invasiediepte zou onderschatten.

Baba (2020) onderzocht aan de hand van retrospectieve data wat de correlatie tussen MRI en histopathologie was bij het meten van invasiediepte in het buccale slijmvlies. Er werd een correlatie gerapporteerd tussen coronale T2-gewogen MRI en histopathologie (Spearman's r=0.67, p=0.012) en tussen coronale T1-gewogen MRI met contrast en vetsupressie (CET1) en histopathologie (Spearman's r=0.68, p<0.001). De auteurs concludeerden dat MRI-metingen behulpzaam zouden kunnen zijn bij het schatten van de histopathologische invasiediepte.

Chin (2020) rapporteerde de overeenkomst tussen contrast CT en histopathologie voor het meten van invasiediepte bij patiënten met plaveiselcelcarcinomen op de tong. Voor de overeenstemming tussen axiale contrast CT en histopathologie werd een ICC (ICC=0.96, 95%BHI: 0.89-0.98, p=<0.001) en een Bland-Altman plot (gemiddeld verschil: -0.72 millimeter,95% limieten van overeenstemming: 3.34 tot -4.78 millimeter) gerapporteerd. Voor een coronale contrast CT was de ICC 0.957 (95%BHI: 0.86-0.99, p<0.001) en het gemiddelde verschil -1.11 millimeter (95% limieten van overeenstemming: 2.73 tot -4.93 millimeter). De auteurs concludeerden dat er een excellente overeenstemming was tussen contrast CT en histopathologie.

Cocker (2020) gebruikte retrospectieve data om de overeenstemming tussen histopathologie en verschillende modaliteiten voor het meten van invasiediepte bij patiënten met mondholtecarcinomen. Er werd data voor echografie (exacte overeenstemming: 9 metingen, binnen 3 millimeter: 52 metingen, buiten 3 millimeter: 17 metingen), MRI (exacte overeenstemming: 1 meting, binnen 3 millimeter: 58 metingen, buiten 3 millimeter: 45 metingen) en CT (exacte overeenstemming: 1 metingen, binnen 3 millimeter: 11 metingen, buiten 3 millimeter: 9 metingen) gerapporteerd. De auteurs concludeerden dat, van de drie modaliteiten, echografie de meest betrouwbare modaliteit was en dat de huidige beeldvormende modaliteiten wellicht geen robuuste en accurate metingen geven.

Filauro (2020) rapporteerde de correlatie van MRI of echografie met histopathologie voor metingen van invasiediepte bij patiënten met mondholtecarcinomen. Er werd een correlatie gevonden tussen metingen met MRI en histopathologie (Spearman's r=0.83, p<0.000) en tussen echografie en histopathologie (Spearman's r=0.76, p<0.0001). De overeenstemming op T-stadium werd tevens vermeld voor MRI en histopathologie (gewogen Kappa=0.53, 95%BHI: 0.32-0.74, p<0.0001) en voor echografie en histopathologie (gewogen Kappa=0.36, 95%BHI: 0.14-0.58). De auteurs concludeerden dat beide modaliteiten valide manieren zijn om preoperatief invasiediepte te bepalen, hetzij met andere kosten-effectiviteitsprofielen en indicaties.

Waech (2020) onderzocht de correlatie tussen MRI of contrast CT en histopathologie voor het meten van invasiediepte bij patiënten met mondholtecarcinomen. De Spearman's r werd gerapporteerd voor de correlatie van T1-gewogen MRI (r=0.635, p<0.0001), T2-gewogen MRI (r=0.679, p<0.0001) of contrast CT (r=0.718, p<0.0001) met histopathologie. De auteurs concludeerden dat preoperatieve metingen met MRI of CT tot een overschatting van de histologische invasiediepte leiden, vooral in tumoren met een invasiediepte kleiner dan vijf millimeter.

Wanneer invasiediepte gebruikt wordt voor T-stadiering kan door over- of onderschatting het klinische T-stadium veranderen. Aangezien het T-stadium op basis van invasiediepte meestal niet het primaire behandelplan beïnvloedt, lijkt dit minder belangrijk voor T-stadiering dan het bepalen van de resectie. Indien het beleid ten aanzien van het electief behandelen van de lymfeklieren in de hals gebaseerd is op invasiediepte (in plaats van schildwachtklierprocedure of beeldvormend onderzoek van de hals) kan onder- of overschatting wel het beleid van de hals beïnvloeden.

Voordeel van een MRI is dat deze meestal al gemaakt wordt om de uitbreiding van de primaire tumor en lymfeklieren in de hals te onderzoeken. Ook kunnen later de invasiediepte en tumordikte nog bepaald worden. Er kunnen contra-indicaties zijn voor het gebruik van een MRI. Dit betreffen de gebruikelijke contra-indicaties voor MRI. Patiënten dienen, zoals gangbaar is, gescreend te worden op deze contra-indicaties. Patiënten zouden eventueel een andere beeldvorming kunnen prefereren wegens claustrofobische klachten die bij het afnemen van een MRI zouden kunnen optreden. Echografie is doorgaans meer afhankelijk van de ervaring van de verrichter en metingen kunnen alleen real-time verricht worden. Daarbij moet de tumor voldoende bereikbaar zijn voor de intraorale probe. Hoewel met de zogenaamde hockey-stick probes tumoren ook achter in de mond beter bereikbaar zijn geworden, kunnen factoren als de (veranderde) lokale anatomie, dentitie, trismus en tumorgrootte een goede meting verhinderen. Voor het gebruik van intra-orale echografie zijn er geen directe contra-indicaties, behalve wanneer de tumor niet te bereiken is. Hierbij is, bijvoorbeeld, te denken aan pijnklachten, trismus en/of de tumorlocatie die het bereiken van de tumor met de probe niet mogelijk maken. Deze echografie kan gemakkelijk tegelijk met een eventuele echografie van de lymfeklieren in de hals plaatsvinden. Hiervoor dient dan alleen de probe gewisseld te worden.

In de diagnostische work-up worden er al vaak een MRI en/of echografie uitgevoerd voor andere doeleinden. Deze beelden zijn echter ook te gebruiken voor het bepalen van de invasiediepte (en/of tumordikte), zonder dat er nieuwe beeldvorming plaats hoeft te vinden. Daardoor zullen er weinig extra kosten worden verwacht bij het gebruik van MRI of intra-orale echografie. Wanneer met verwacht dat een oppervlakkige tumor niet zichtbaar is op een MRI kan echografie overwogen worden indien de locatie bereikbaar is. De lokalisatie van de tumor kan leidend zijn voor de keuze tussen MRI of echografie. Vooral voor tumoren op de tong is er data over het gebruik van echografie bekend, maar minder voor andere tumorlocaties. Voor het gebruik van de intra-orale echografie kan, specifiek bij moeilijk bereikbare tumorlocaties, een hockey-stick probe worden gebruikt. De werkgroep beseft zich dat deze probe niet overal beschikbaar is en dat aanschaf van een dergelijke probe extra kosten met zich meebrengt. Ook de extra tijd die echografie in beslag neemt en training van personeel kunnen kosten met zich meebrengen. Zowel MRI als echografie zijn geaccepteerde beeldvormingstechnieken in de diagnostische work-up van patiënten met hoofd-halstumoren, waardoor er geen problemen worden verwacht in de aanvaardbaarheid, haalbaarheid en implementatie.

Aanbeveling

Aanbeveling-1

<u>Rationale van de aanbeveling: weging van argumenten voor en tegen de interventies</u>

Beeldvormingstechnieken, zoals MRI en intra-orale echografie, die voor andere doeleinden worden ingezet bij patiënten met hoofd-halstumoren kunnen worden gebruikt om de invasiediepte te meten zonder grote bijkomende kosten. MRI en intra-orale echografie lijken de voorkeur te hebben boven de mogelijke alternatieven, zoals palpatie of CT. Doordat er weinig informatie beschikbaar is en er weinig zekerheid bestaat over het gebruik van

palpatie, acht de werkgroep het van belang om in elk geval een beeldvormingstechniek te gebruiken voor het meten van de invasiediepte.

Bepaal de invasiediepte met een beeldvormingstechniek, bij voorkeur MRI, met als alternatief intraorale echografie.

Literatuur

Alsaffar HA, Goldstein DP, King EV, de Almeida JR, Brown DH, Gilbert RW, Gullane PJ, Espin-Garcia O, Xu W, Irish JC. Correlation between clinical and MRI assessment of depth of invasion in oral tongue squamous cell carcinoma. J Otolaryngol Head Neck Surg. 2016 Nov 22;45(1):61. PubMed PMID: 27876067; PubMed Central PMCID: PMC5120480.

Baba A, Masuda K, Hashimoto K, Matsushima S, Yamauchi H, Ikeda K, Yamazaki M, Suzuki T, Ogane S, Kurokawa R, Kurokawa M, Ota Y, Mogami T, Nomura T, Ojiri H. Correlation between the magnetic resonance imaging features of squamous cell carcinoma of the buccal mucosa and pathologic depth of invasion. Oral Surg Oral Med Oral Pathol Oral Radiol. 2021 Jan 8:S2212-4403(21)00002-X. doi: 10.1016/j.oooo.2020.12.023. Epub ahead of print. PMID: 33516643.Brouwer de Koning SG, Karakullukcu MB, Lange CAH, Ruers TJM. The oral cavity tumor thickness: Measurement accuracy and consequences for tumor staging. Eur J Surg Oncol. 2019 Nov;45(11):2131-2136. doi: 10.1016/j.ejso.2019.06.005. Epub 2019 Jun 4. PubMed PMID: 31227341.

Chin SY, Kadir K, Ibrahim N, Rahmat K. Correlation and accuracy of contrast-enhanced computed tomography in assessing depth of invasion of oral tongue carcinoma. Int J Oral Maxillofac Surg. 2020 Nov 5:S0901-5027(20)30377-5. doi: 10.1016/j.ijom.2020.09.025. Epub ahead of print. PMID: 33162298.

Choi N, Noh Y, Lee EK, Chung M, Baek CH, Baek KH, Jeong HS. Discrepancy between cTNM and pTNM staging of oral cavity cancers and its prognostic significance. J Surg Oncol. 2017 Jun;115(8):1011-1018. doi: 10.1002/jso.24606. Epub 2017 Mar 23. PubMed PMID: 28334428.

Cocker H, Francies O, Adams A, Sassoon I, Schilling C. Do we have a robust method for preoperative tumour depth assessment for oral cavity tumours with clinically negative necks? Int J Oral Maxillofac Surg. 2020 Dec 25:S0901-5027(20)30416-1. doi: 10.1016/j.ijom.2020.11.002. Epub ahead of print. PMID: 33358587.

Filauro M, Missale F, Marchi F, Iandelli A, Carobbio ALC, Mazzola F, Parrinello G, Barabino E, Cittadini G, Farina D, Piazza C, Peretti G. Intraoral ultrasonography in the assessment of DOI in oral cavity squamous cell carcinoma: a comparison with magnetic resonance and histopathology. Eur Arch Otorhinolaryngol. 2020 Oct 21. doi: 10.1007/s00405-020-06421-w. Epub ahead of print. PMID: 33084951.

Goel V, Parihar PS, Parihar A, Goel AK, Waghwani K, Gupta R, Bhutekar U. Accuracy of MRI in Prediction of Tumour Thickness and Nodal Stage in Oral Tongue and Gingivobuccal Cancer With Clinical Correlation and Staging. J Clin Diagn Res. 2016 Jun;10(6):TC01-5. doi: 10.7860/JCDR/2016/17411.7905. Epub 2016 Jun 1. PubMed PMID: 27504375; PubMed Central PMCID: PMC4963735.

Iida Y, Kamijo T, Kusafuka K, Omae K, Nishiya Y, Hamaguchi N, Morita K, Onitsuka T. Depth of invasion in superficial oral tongue carcinoma quantified using intraoral ultrasonography. Laryngoscope. 2018 Dec;128(12):2778-2782. doi: 10.1002/lary.27305. Epub 2018 Oct 16. PubMed PMID: 30325049.

Klein Nulent TJW, Noorlag R, Van Cann EM, Pameijer FA, Willems SM, Yesuratnam A, Rosenberg AJWP, de Bree R, van Es RJJ. Intraoral ultrasonography to measure tumor thickness of oral cancer: A systematic review and meta-analysis. Oral Oncol. 2018 Feb;77:29-36. doi: 10.1016/j.oraloncology.2017.12.007. Epub 2017 Dec 18. PubMed PMID: 29362123.

Mao MH, Wang S, Feng ZE, Li JZ, Li H, Qin LZ, Han ZX. Accuracy of magnetic resonance imaging in evaluating the depth of invasion of tongue cancer. A prospective cohort study. Oral Oncol. 2019 Apr;91:79-84. doi: 10.1016/j.oraloncology.2019.01.021. Epub 2019 Mar 4. PubMed PMID: 30926067.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., … de Vet, H. C. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation, 19(4), 539-49.

Mokkink LB, Prinsen CA, Patrick DL, Alonso J, Bouter LM, de Vet HC, Terwee CB, Mokkink LB. COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs). User manual. 2018;78:1. Available from: https://www.cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018-1.pdf.

Nair AV, Meera M, Rajamma BM, Anirudh S, Nazer PK, Ramachandran PV. Preoperative ultrasonography for tumor thickness evaluation in guiding management in patients with early oral tongue squamous cell carcinoma. Indian J Radiol Imaging. 2018 Apr-Jun;28(2):140-145. doi: 10.4103/ijri.IJRI_151_17. PubMed PMID: 30050234; PubMed Central PMCID: PMC6038222.

Noorlag R, Klein Nulent TJW, Delwel VEJ, Pameijer FA, Willems SM, de Bree R, van Es RJJ. Assessment of tumour depth in early tongue cancer: Accuracy of MRI and intraoral ultrasound. Oral Oncol. 2020 Jul 9;110:104895. doi: 10.1016/j.oraloncology.2020.104895. Epub ahead of print. PMID: 32653839.

Prinsen CA, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, et al. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" – a practical guideline. Trials. 2016;17(1):449.

Shintani S, Yoshihama Y, Ueyama Y, Terakado N, Kamei S, Fijimoto Y, Hasegawa Y, Matsuura H, Matsumura T. The usefulness of intraoral ultrasonography in the evaluation of oral cancer. Int J Oral Maxillofac Surg. 2001 Apr;30(2):139-43. PubMed PMID: 11405449.

Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34-42.

Verma, A., Singhal, A., Hadi, R., Singh, P. and Raj, G., 2019. Evaluation of tumor thickness in three dimensions on magnetic resonance imaging and its comparison with final histopathology in squamous cell carcinoma of the tongue. *Clinical Cancer Investigation Journal*, *8*(4), p.161.

Vidiri A, Panfili M, Boellis A, Cristalli G, Gangemi E, Pellini R, Marzi S, Covello R. The role of MRI-derived depth of invasion in staging oral tongue squamous cell carcinoma: inter-reader and radiological-pathological agreement. Acta Radiol. 2019 Jul 18:284185119862946. doi: 10.1177/0284185119862946. (Epub ahead of print) PubMed PMID: 31319692.

Waech T, Pazahr S, Guarda V, Rupp NJ, Broglie MA, Morand GB. Measurement variations of MRI and CT in the assessment of tumor depth of invasion in oral cancer: A retrospective study. Eur J Radiol. 2021 Feb;135:109480. doi: 10.1016/j.ejrad.2020.109480. Epub 2020 Dec 15. PMID: 33370639.

Geldigheid en Onderhoud

Module[1]	Regiehouder(s)[2]	Jaar van autorisatie	Eerstvolgende beoordeling	Frequentie van	Wie houdt er toezicht	Relevante factoren voor

[1] Naam van de module

[2] Regiehouder van de module (deze kan verschillen per module en kan ook verdeeld zijn over meerdere regiehouders)

			actualiteit richtlijn[3]	beoordeling op actualiteit[4]	op actualiteit[5]	wijzigingen in aanbeveling[6]

[3] Maximaal na vijf jaar
[4] (half)Jaarlijks, eens in twee jaar, eens in vijf jaar
[5] regievoerende vereniging, gedeelde regievoerende verenigingen, of (multidisciplinaire) werkgroep die in stand blijft
[6] Lopend onderzoek, wijzigingen in vergoeding/organisatie, beschikbaarheid nieuwe middelen

Bijlagen bij module 1.1

Kennislacunes

What is the agreement between preoperative clinical examination (by palpation), computed tomography (CT), positron emission tomography/computed tomography (PET-CT), magnetic resonance imaging (MRI) or intraoral ultrasound, and postoperative histopathologic results for measuring the depth of the invasion (or tumor thickness) by a tumor in patients with an oral cavity carcinoma?

Indicatoren

Geen.

Implementatieplan

Aanbeveling	Tijdspad voor implementatie: < 1 jaar, 1 tot 3 jaar of 3 tot 5 jaar	Verwacht effect op kosten	Randvoorwaarden voor implementatie (binnen aangegeven tijdspad)	Mogelijke barrières voor implementatie[1]	Te ondernemen acties voor implementatie[2]	Verantwoordelijken voor acties[3]	Overige opmerkingen
Bepaal de invasiediepte met een beeldvormingstechniek, bij voorkeur MRI, met als alternatief intraorale echografie.	< 1 jaar	MRI wordt in meeste centra al standaard vervaardigd bij mondholtecarcinomen. Bij de meeste patiënten met een mondholtecarcinoom wordt standaard echografie van de hals verricht. Aansluitend kan gemakkelijk een echo van de tumor verricht worden. De meerkosten zitten voornamelijk in de kosten van de speciale echo probe voor intraorale echografie.	MRI dient bij alle patiënten vervaardigd te worden. Beschikbaarheid probe voor intraorale toepassing. Komt binnen landelijke studie naar echogeleide resecties van tongtumoren voor meeste centra beschikbaar.	Bij zeer oppervlakkige tumoren wordt niet altijd een MRI vervaardigd. Niet in alle centra wordt standaard echografie van de hals verricht. Beschikbaarheid echo probe voor intraoraal toepassing.	Indien geen MRI wordt vervaardigd, invasiediepte met echografie bepalen. Beschikbaar maken van intraorale echo probe	Lokale werkgroepen	Vergoeding intraorale echografie niet duidelijk of geregeld is.

[1] Barrières kunnen zich bevinden op het niveau van de professional, op het niveau van de organisatie (het ziekenhuis) of op het niveau van het systeem (buiten het ziekenhuis). Denk bijvoorbeeld aan onenigheid in het land met betrekking tot de aanbeveling, onvoldoende motivatie of kennis bij de specialist, onvoldoende

faciliteiten of personeel, nodige concentratie van zorg, kosten, slechte samenwerking tussen disciplines, nodige taakherschikking, et cetera.
[2] Denk aan acties die noodzakelijk zijn voor implementatie, maar ook acties die mogelijk zijn om de implementatie te bevorderen. Denk bijvoorbeeld aan controleren aanbeveling tijdens kwaliteitsvisitatie, publicatie van de richtlijn, ontwikkelen van implementatietools, informeren van ziekenhuisbestuurders, regelen van goede vergoeding voor een bepaald type behandeling, maken van samenwerkingsafspraken.
[3] Wie de verantwoordelijkheden draagt voor implementatie van de aanbevelingen, zal tevens afhankelijk zijn van het niveau waarop zich barrières bevinden. Barrières op het niveau van de professional zullen vaak opgelost moeten worden door de beroepsvereniging. Barrières op het niveau van de organisatie zullen vaak onder verantwoordelijkheid van de ziekenhuisbestuurders vallen. Bij het oplossen van barrières op het niveau van het systeem zijn ook andere partijen, zoals de NZA en zorgverzekeraars, van belang.

Evidence tables

Study reference	Study characteristics	Patient characteristics	Measurement properties and procedures	Follow-up	Interpretability of results	Outcome measures and effect size [4]	Comments
Klein Nulent 2018 (systematic review)	Instruments assessed: Ultrasound (intraoral) versus. histopathology for **tumor thickness** Included studies: A: Joshi 2014 B: Yesuratnam 2014 C: Chammas 2011 D: Lodder 2011 E: Kodama 2010 F: Mark Taylor 2010 G: Kaneoya 2009 H: Baek 2008 I: Yamane 2007 J: Songra 2006 K: Helbig 2005 L: Shintani 2001 Setting and country: A: India B: Australia C: Brazil D: Netherlands E: Japan F: Canada G: Japan H: Korea I: Japan J: UK K: Germany L: Japan	<u>Inclusion criteria</u>: Patients with OSCC, pre- or intraoperative measurement of tumor thickness or margin assessment by ultrasound, measurements compared to histopathological tumor thickness or margin width. <u>Exclusion criteria</u>: Duplicates, reviews / book chapters/ case reports / editorials / oral presentations / notes / poster presentations, analyses of head and neck SCC without subgroup analysis for oral SCC, languages other than English or German. <u>Search date:</u> 6 july 2016	Describe the assessed measurement properties and their procedures: <u>Reliability</u> *Measurement error (inter-instrument)* Studies included in Bland-altman plot: A: Yes B: Yes C: Yes D: Yes E: Yes F: Yes G: No H: Yes I: No J: Yes K: Yes L: Yes *Timing of US:* A: preoperative B: preoperative C: preoperative D: preoperative E: intraoperative F: preoperative G: unclear H: intraoperative I: unclear J: intraoperative	<u>Incomplete outcome data</u>: No participants where no individual data was available, n: A: all data available B: 22 C: all data available D: 32 E: all data available F: all data available G: not included in bland-altman plot H: all data available I: not included in bland-altman plot. J: all data available K: all data available L: 1 <u>How were missing data handled?</u> Excluded from bland-altman plot. <u>Length of follow-up (if applicable)</u>: Not reported in the SR, somewhat deducible from	Was the distribution of the (total) scores in the study sample described? (yes/no): Not described Percentage of the sample with the lowest score possible: Not relevant Percentage of the sample with the highest score possible: Not relevant Minimally important change/difference determined or referred? (yes/no) Not relevant	Outcome measures and effect size (include 95%CI and p-value if available): <u>Reliability</u> *Measurement error* Bland-Altman plot (US-histopathology), mm: Mean difference: 0.5. 95% upper limit: 6.5 95% lower limit: -5.5	Authors used the QUADAS-2 tool for the risk of bias assessment. Individual patient data from Taylor 2010 was estimated from a figure by the SR authors. TNM edition: TNM7 (data was extracted according to AJCC7)

	Funding and conflicts of interest: SR authors declare that they have no conflicts of interest and no specific grants were reveived. CoI and funding is not reported for included studies	<u>Search sources:</u> PubMed, Embase, Cochrane databases <u>Sample characteristics[1]</u>: *Sample size (in Bland-altmanplot), n:* A: 7 B: 66 C: 19 D: 33 E: 13 F: 21 G: - H: 20 I: - J: 14 K: 9 L: 38 *Tumor site:* A: tongue and buccal mucosa B: tongue C: tongue D: tongue and FOM E: tongue F: tongue and FOM G: tongue H: tongue I: tongue J: tongue, FOM, lip, alveolar mucosa K: tongue	K: intraoperative L: unclear *Device and probe:* A: RIC5 GE Voluson (9-5MHz) B: Philips iU22 (7-15MHz) C: GE Med. Systems Logiq 500 (5-10MHz) D: Philips iU22 (5-7MHz, 7-15MHz) E: Aloka SSD-1200CV (7.5MHz) F: unclear (10-12MHz) G: Toshiba PLM-1202S (12MHz) H: Aloka UST-9120 (8-10MHz) I: Aloka SSD-630 (10MHz) J: AT HDI-5000 (5-10MHz) K: Diasonics VERSUST (8-12MHz) L: Toshiba PEF-704LA (7.5MHz), Aloka UST-995 (7.5MHz), Aloka UST-5536 (7.5MHz) Preoperative or intraoperative ultrasound was compared to a histopathological assessment.	'Flow and timing' in the risk of bias assessment; 9/16 unclear, 1/16 high risk, 6/16 low risk. <u>Loss-to-follow-up (if applicable)</u>: NA			

		L: tongue, FOM, buccal mucosa					
Alsaffar 2016	Instruments assessed: Clinical assessment versus histopathology and MRI assessment versus histopathology for **depth of invasion** (categorized) Setting and country: Cancer center, Canada Sampling method: Patients referred to the cancer center were recruited. Funding and conflicts of interest: Authors had no CoI or funding	Inclusion criteria: Newly diagnosed oral SCC Exclusion criteria: Referred with an MRI that was reviewed by the surgeon prior to clinical exam and enrolment in the study, CT imaging only, carcinoma in situ or previous excisional biopsy, previous head and neck (chemo)radiation. N total at baseline: N= 53 Sample characteristics[1]: *Mean age ± SD (or median age (range)):* *64* *Sex (male/female): G: 34/19* *T-stage, n:* T1: 22 T2: 22	Describe the assessed measurement properties and their procedures: Reliability *Measurement error* Clinical examination consisted of the treating surgeon performing a palpation prior to radiographic evaluation of the tumor depth. For MRI and pathologic assessment, the depth of invasion was measured form the adjacent mucosa to the deepest tumor aspect. Invasion depth was categorized (binominal) using 2 categories: < 5mm invasion, and ≥5mm invasion.	Incomplete outcome data: From sample (and subgroups if applicable) NA How were missing data handled? NA Length of follow-up (if applicable): Unclear time between preoperative assessment and pathological assessment. Loss-to-follow-up (if applicable): NA	Was the distribution of the (total) scores in the study sample described? Pathological depth <5: 10 Pathological depth ≥5: 43 Radiological depth <5: 9 Radiological depth ≥5: 40 Percentage of the sample with the lowest score possible: Not relevant Percentage of the sample with the highest score possible: Not relevant Minimally important change/difference determined or referred? (yes/no) Not relevant	Outcome measures and effect size (include 95%CI and p-value if available): Reliability *Measurement error* Kappa at a cut-off point of 5mm invasion depth for MRI: K=0.80 (95%CI: 0.59-1.00) Kappa at a cut-off point of 5mm invasion depth for clinical assessment: K=0.61 (95%CI: 0.36-0.87)	Unclear time between preoperative assessment and pathological assessment. TNM edition: unclear

		T3: 7 T4: 2 *N-stage, n:* N0: 32 N1: 7 N2: 11 N3: 0					
Brouwer de Koning, 2019	Instruments assessed: Ultrasound versus histopathology and MRI versus histopathology for **tumor thickness** Setting and country: Hospital, Netherlands Sampling method: Database between 2011 and 2016 Funding and conflicts of interest: Authors declare that they have no CoI. Funding not reported.	Inclusion criteria: Oral cavity cancer that was clinically staged as T1 or 2, US or MR images were acquired between 2011-2016. Exclusion criteria: None reported. N total at baseline: N=142 treated, n=83 included Sample characteristics[1]: *Mean age ± SD (or median age (range)):* *61 (range: 31-88)* *Sex (male/female):* *G: 45M/38F* *Disease site, n:* Tongue: 58 FOM: 24	Describe the assessed measurement properties and their procedures: Reliability *Measurement error* Tumor thicknes was measured in n=76 with US and in n=46 with MRI. Preoperative MRI and US were performed (US after excision was excluded). For US, a Hitachi EUB-900 was used, with the EUP-054 transducer at 13-7MHz. The probe was placed directly on the lesion. MRI was performed with a Philips Achieva 3T (dedicated 16-channel SENSE neurovascular coil). Used sequences: T1W, TSE, TRA, TR, TE	Incomplete outcome data: From sample (and subgroups if applicable) N=32 was excluded because tumor dimensions were reported by only 1 modality; n=11 was excluded because no histopathology data was available; n=5 were excluded because US was acquired after excision, n=1 because histopathology showed scar tissue, n=10 tumor was not assessable with US. (leaving n=83 for analysis) How were missing data handled?	Was the distribution of the (total) scores in the study sample described? US tumor thickness, mean (SD) in mm: 5.1 (3.1) Histopathology tumor thickness, mean (SD) in mm: 5.1 (3.5) MR tumor thickness, mean (SD) in mm: 7.4 (3.5) Histopathology tumor thickness, mean (SD) in mm: 6.1 (3.2) Percentage of the sample with the lowest score possible: Not relevant Percentage of the sample with the highest score possible: Not relevant	Outcome measures and effect size (include 95%CI and p-value if available): Reliability *Measurement error* Bland-Altman plot (US-histopathology), mm: Mean difference: 0.05. 95% upper limit: 5.4 95% lower limit: -5.3 Bland-Altman plot (MRI-histopathology), mm: Mean difference: 1.3. 95% upper limit: 8.6 95% lower limit: -6.1	All tumors were SCC. TNM edition: AJCC7 and AJCC8 were used. However, the AJCC7 was used to classify in T-stage (using the greatest dimension)

		Palate: 2 Lip: 1	538/10ms, flip angle 90, matrix 288/248, slice thickness of 4mm, STIR TSE COR, TR/TE 2500/60ms, matrix 216/170, T1 3D Thrive fat-saturation, intravenous injection of 15cc gadoterate meglumine, TR/TE 9.86/4.59ms, flip angle 10, matrix 200/179, slice thickness 1mm. The tumor dimensions were measured in 3D and were reported by the radiologist (suggesting a T-stage). Reports of the radiologist were used for the study. Tumor dimensions were also reported by the pathologist in the pathological report. Further procedures on the histopathological measurement were not reported.	Patients were excluded from analysis. <u>Length of follow-up (if applicable)</u>: Unclear time between preoperative assessment and histopathology. <u>Loss-to-follow-up (if applicable)</u>: NA	Minimally important change/difference determined or referred? (yes/no) Not relevant		
Choi 2017	Instruments assessed: composite clinical assessment (endoscopic + palpation + CT or MRI) versus histopathology for **tumor thickness** (categorized for T-stage)	<u>Inclusion criteria</u>: Biopsy-proven SCC of the oral cavity had undergone curative surgical resection of primary tumor and neck dissection or	Describe the assessed measurement properties and their procedures: <u>Reliability</u> *Measurement error*	<u>Incomplete outcome data</u>: From sample (and subgroups if applicable) NA	Was the distribution of the (total) scores in the study sample described? (yes/no): *pT classification, n (%):* T1: 109 (43.3%) T2: 80 (31.7%) T3: 25 (9.9%)	Outcome measures and effect size (include 95%CI and p-value if available): <u>Reliability</u> *Measurement error*	Unclear time period between preoperative and postoperative assessment.

	Setting and country: Hospital, Korea Sampling method: Database between 1996 and 2012 Funding and conflicts of interest: Funded by Samsung biomedical research institute basic clinical collaborative research grant and National Research Foundation of Korea. Authors state that there were no CoI to declare.	sentinel node biopsy as initial treatment. Exclusion criteria: Synchronous or metachronous cancers, distant metastasis N total at baseline: N=252 Sample characteristics[1]: *Median age range: 55(range: 47-65)* *Sex (male/female): 164M/88F* *Tumor site, n (%):* Tongue: 195 (77.4%) FOM: 34 (13.5%) Buccal: 23 (9.1%) *pT classification, n (%):* T1: 109 (43.3%) T2: 80 (31.7%) T3: 25 (9.9%) T4a: 37 (14.7%) T4b: 1 (0.4%)	Clinical staging was performed preoperatively by physicians with endoscopes (magnified view) and palpation and either contrast-enhanced CT or MRI. No further procedures described. Surgical specimens of the primary tumor were assessed grossly and microscopically. No further procedures described.	How were missing data handled? NA Length of follow-up (if applicable): Unclear time period between preoperative assessment and postoperative histopathological assessment Loss-to-follow-up (if applicable): From sample (and subgroups if applicable) NA	T4a: 37 (14.7%) T4b: 1 (0.4%) *cT classification, n (%):* T1: 114 (45.2%) T2: 87 (34.5%) T3: 25 (9.9%) T4a: 37 (14.7%) T4b: 1 (0.4%) Percentage of the sample with the lowest score possible: NA Percentage of the sample with the highest score possible: NA Minimally important change/difference determined or referred? (yes/no) NA	Kappa for the agreement of T-stage (clinical examination versus pathological data): K=0.81 (95%CI: not reported, p<0.001)	TNM edition: TNM7 (AJCC7) was used.
Goel 2016	Instruments assessed: clinical versus histopathology, and MRI	Inclusion criteria: Biopsy proven SCC of the tongue or	Describe the assessed measurement properties and their procedures:	Incomplete outcome data:	Was the distribution of the (total) scores in	Outcome measures and effect size (include	Agreement was categorized in T-stages.

	versus histopathology for **tumor thickness** (categorized for T-stage) Setting and country: Hospital, India Sampling method: Prospective data-gathering of consecutive participants between July 2013 and august 2015. Funding and conflicts of interest: authors state that there are no financial or competing interests.	gingiva-buccal region with enlarged neck nodes <u>Exclusion criteria</u>: Claustrophobic patients, metallic implants, other cancers of oral cavity, MR staging > T4a, patients not willing to undergo MRI. <u>N total at baseline</u>: N=61 <u>Sample characteristics[1]</u>: *Mean age ± SD (or median age (range)):* *Sex (male/female):* *G: 45M/16F* *T-stage, n:* T1: 4 T2: 16 T3: 13 T4: 28	<u>Reliability</u> *Measurement error* No procedures of the clinical examination were reported. MRI was performed with 1.5-T (BRIVO MR 355 1.5Tesla GE MRI). Patients lay supine on the MRI table and a head coil was applied. Sequences of 4mm thickness with 1mm intersection gap and a 256x256 matrix were used (240mm FOV). Routine T1WI, T2WI, Coronal STIR followed by post contrast axial T1W were performed. Axial, coronal T2WI and post contrast T1WI were used to measure the lesion size. Lesions were stages (T-stage). No procedures for histopathology were reported. The agreement of clinical examinarion or MRI with	From sample (and subgroups if applicable) NA <u>How were missing data handled?</u> NA <u>Length of follow-up (if applicable)</u>: Unclear how much time there was between clinical or MRI examination and histopathological assessment. <u>Loss-to-follow-up (if applicable)</u>: From sample (and subgroups if applicable) NA	the study sample described? (yes/no): *T-stage histopathology, n:* T1: 4 T2: 16 T3: 13 T4: 28 *T-stage clinical examination, n:* T1: 0 T2: 34 T3: 11 T4: 16 *T-stage MRI, n:* T1: 5 T2: 18 T3: 9 T4: 29 Percentage of the sample with the lowest score possible: Not relevant Percentage of the sample with the highest score possible: Not relevant Minimally important change/difference determined or referred? (yes/no)	95%CI and p-value if available): <u>Reliability</u> *Measurement error* Kappa for the agreement of T-stage (clinical examination versus pathological data): K=0.47 (95%CI: not reported) Kappa for the agreement of T-stage (MRI versus pathological data): K=0.69 (95%CI: not reported)	No procedures for histopathology were reported. TNM edition: Unclear

			histopathology on T-stage was assessed (T1-T4).		Not relevant		
Iida 2018	Instruments assessed: Ultrasound versus histopathology for **tumor invasion** (categorized) Setting and country: US was performed in an outpatient clinic, Hospital, Japan. Sampling method: Funding and conflicts of interest:	<u>Inclusion criteria</u>: Early oral tongue SCC, patient between June 2008 and December 2015, <u>Exclusion criteria</u>: Local recurrence after partial glossectomy, prior radiotherapy, prior chemotherapy. <u>N total at baseline</u>: N=56 <u>Sample characteristics</u>[1]: *Mean age ± SD):* *59 (range: 25-90)* *Sex (male/female):* *G: 34M/22F* *Disease site, n:* Lateral ledge: 56 Other: 0 *Tumor size by clinical measurement, n:* *≤20mm: 44* *>20mm: 12*	Describe the assessed measurement properties and their procedures: <u>Reliability</u> *Measurement error* Histological assessment was performed by using a micrometer in the tumor specimen (formalin-fixed paraffin-embedded). Depth of invasion was measured from the level of basement membrane of the closest normal mucosa. Proeoperative intraoral US was performed in an outpatient clinic. A Hitachi Aloka UST-5713T (16MHz) was used. The patient extended the tongue, which was held with gauze on the contralateral side. The reference line was defined as the line connecting tumor-normal mucosal junction of both sides.	<u>Incomplete outcome data</u>: NA <u>How were missing data handled?</u> NA <u>Length of follow-up (if applicable)</u>: The time between preoperative measurement and histopathological assessment was unclear. <u>Loss-to-follow-up (if applicable)</u>: NA	Was the distribution of the (total) scores in the study sample described? Invasion assessed by histology, median (range): 3.5 (0.0-12.0) Invasion assessed by ultrasound, median (range): 3.6 (0.7-9.2) Percentage of the sample with the lowest score possible: Not relevant Percentage of the sample with the highest score possible: Not relevant Minimally important change/difference determined or referred? (yes/no) Not relevant	Outcome measures and effect size (include 95%CI and p-value if available): <u>Reliability</u> *Measurement error* Kappa at a cut-off point of 5mm invasion depth: K=0.651 (95%CI: 0.43-0.87)	Instrument agreement was categorized using a cutoff value of 5mm, resulting in 2 categories: ≤5mm invasion, and >5mm invasion. TNM edition: Unclear

			Measurement had a resolution of 0.1mm. Instrument agreement was categorized using a cutoff value of 5mm, resulting in 2 categories: ≥5mm invasion, and <5mm invasion.				
Mao 2019	Instruments assessed: MRI versus histopathology for **depth of invasion** Setting and country: Hospital, China Sampling method: Prospective, from April 2015 to December 2017 Funding and conflicts of interest: funding by Discipline Construction Fund of Beijing Stomatology Hospital (grant/award number 17-09-14). The authors declared that the funding source had no role in the study. The authord declared that there was no CoI.	Inclusion criteria: First diagnosis of SCC of the tongue Exclusion criteria: Ineligible for MRI, T4-stage, recurrent disease, received neoadjuvant treatment, prior radiotherapy. N total at baseline: N=150 Sample characteristics[1]: *Mean age ± SD: 58.01 (SD: 12.10)* *Sex (male/female): G: 80M/70F* *Disease location, n:* Ventral tongue: 35 Tongue border: 89 Dorsal tongue: 19 Tongue base: 7	Describe the assessed measurement properties and their procedures: Reliability *Measurement error* A 1.5T MR was used with section thickness of 1mm (Siemens Magnetom Aera). Proeperative MRI was performed within one week before surgery. The scanning protocol consisted of: T1 axial / coronal / saggital sequences, T2 axial / coronal sequences with fat suppression, T1-weighted axial / coronal / sagittal sequences with fat suppression and contrast media. Intraoperative tumor specimens were obtained by dissecting along the coronal/axial	Incomplete outcome data: From sample (and subgroups if applicable) NA How were missing data handled? G: Length of follow-up (if applicable): Time between preoperative imaging and histopathological assessment was unclear, however preoperative imaging was performed within 1 week prior to surgery. Loss-to-follow-up (if applicable): NA	Was the distribution of the (total) scores in the study sample described? (yes/no): Depth of invasion, mean mm (SD): MRI: 11.75 (6.49) Pathological: 9.43 (5.57) Percentage of the sample with the lowest score possible: Not relevant Percentage of the sample with the highest score possible: Not relevant Minimally important change/difference determined or referred? (yes/no) Not relevant	Outcome measures and effect size (include 95%CI and p-value if available): Reliability *Measurement error* Bland-Altman plot overall n=150 (MRI-histopathology), mm: Mean difference: 2.32. 95% upper limit: 5.61 95% lower limit: -0.97 **T-stage** Bland-Altman plot tumor T1-stage n=43 (MRI-histopathology), mm: Mean difference: 1.48. 95% upper limit: 3.63 95% lower limit: -0.67 Bland-Altman plot tumor T2-stage n=71 (MRI-histopathology), mm: Mean difference: 2.08.	Patient characteristics were extracted from table 1 in the article's supplement. TNM edition: TNM7 (AJCC7)

		Morphology, n: Ulcer: 41 Invasive: 94 Exogeneous: 15 *T-stage, n:* T1: 43 T2: 71 T3: 36 *pN-stage, n:* N1: 16 N2b: 17 N2c: 2	interval of 3mm, and the depth of invasion was measured on a micrometer. Surgical specimens were preserved in formalin. Pathological sections and staining were used to measure the depth of invasion (i.e., the vertical distance between the simulated normal mucosal junction and the deepest point of infiltration).			95% upper limit: 4.62 95% lower limit: -0.45 Bland-Altman plot tumor T3-stage n=36 (MRI-histopathology), mm: Mean difference: 3.79. 95% upper limit: 7.70 95% lower limit: -0.13 **Morphology** Bland-Altman plot ulcer type tumor n=41 (MRI-histopathology), mm: Mean difference: 3.72. 95% upper limit: 7.60 95% lower limit: -0.16 Bland-Altman plot invasive type tumor n=91 (MRI-histopathology), mm: Mean difference: 1.83. 95% upper limit: 4.25 95% lower limit: -0.59 Bland-Altman plot exogeneous type tumor n=15 (MRI-histopathology), mm: Mean difference: 1.53. 95% upper limit: 3.06 95% lower limit: 0.01	

Nair 2018	Instruments assessed: Ultrasound versus histopathology for **tumor thickness** Setting and country: hospital, India Sampling method: Recruitment between January 2012 and December 2013. Unclear recruitment method. Funding and conflicts of interest: authors declare that there was no funding and CoI	<u>Inclusion criteria</u>: Biopsy proven T1N0 or T2N0 primary SCC of the tongue, tumor located on the lateral two-third of the tongue. <u>Exclusion criteria</u>: Tongue tumor crossing the midline of the tongue or involving the tip of the tongue, lateral surface of anterior two-third of the tongue infiltrating into surrounding structures, irradiated tumor, recurrent tumor, tumor of other subsites in the oral cavity. <u>N total at baseline</u>: N=24 <u>Sample characteristics</u>[1]: *Mean age ± SD (or median age (range)):* *55 (range: 22-76)* *Sex (male/female):* *G: 16M/8F*	Describe the assessed measurement properties and their procedures: <u>Reliability</u> Preoperative measurements were performed with ultrasound at 17 or 19MHz. The tongue was extended, and the probe was placed directly on the tumor. Tumor thickness was measures from the surface to the deepest point of invasion. For ulcer type tumors an imaginary line was drawn over the area, joining the normal mucosa on both ends and the deeperst point of invasion was measured. After resection the specimens were placed in saline (not fixed with formalin) and send to the pathology department. The specimen was cut into 2-3mm thick transverse slices.	<u>Incomplete outcome data</u>: From sample (and subgroups if applicable) NA <u>How were missing data handled?</u> NA <u>Length of follow-up (if applicable)</u>: Unclear time between preoperative US assessment and histopathological assessment. <u>Loss-to-follow-up (if applicable)</u>: From sample (and subgroups if applicable) NA	Was the distribution of the (total) scores in the study sample described? (yes/no): Pathological (range): 2-15mm Ultrasound (range): 1-14mm Deduced from a figure (table 3) Percentage of the sample with the lowest score possible: Not relevant Percentage of the sample with the highest score possible: Not relevant Minimally important change/difference determined or referred? (yes/no) Not relevant	Outcome measures and effect size (include 95%CI and p-value if available): <u>Reliability</u> *Reliability* ICC=0.821 (95%CI not provided, ICC model not reported) *Measurement error* Bland-Altman plot overall n=150 (US-histopathology), mm: Mean difference: -0.15 95% upper limit: not reported. 95% lower limit: not reported. LoA's can be approximated from the provided bland-altman plot: UL: 4.99 LL: -4.6	Relatively large spread in histological tumor thickness (2-15mm) may influence the ICC; Pearson's r=0.69, ICC=0.82. 95% LoA's were approximated from the bland-altman plot (table 4). TNM edition TNM6 (AJCC6)

		T-stane, n: pT1: 18 pT2: 6					
Shintani 2001	Instruments assessed: CT versus histopathology and MRI versus histopathology for **tumor thickness** Setting and country: Hospital, Japan Sampling method: Unclear Funding and conflicts of interest: Unclear	Inclusion criteria: Unclear Exclusion criteria: Unclear N total at baseline: N=38 (38 had CT scans and 24 also had an MRI) Sample characteristics[1]: *Mean age ± SD: 58.2 (SD not reported, range: 36-91)* *Sex (male/female): Not reported* *Tumor location, n:* Tongue: 26 Buccal: 8 FOM: 4	Describe the assessed measurement properties and their procedures: Reliability *Measurement error* Tumor thickness was measured by a contrast-enhanced 5-mm axials CT scan in 39 participants. A 4-mm axial and coronal T1/T2-weighted MRI was used in 26 participants. An ocular micrometer was used for histological sections.	Incomplete outcome data: Study included n=39, but n=38 was reported in table 1. MRI was not available in n=14 participants. Small tumors were not detected by CT and MRI: 19/38 not detected by CT, 11/24 not detected by MRI. How were missing data handled? Excluded from analysis. Length of follow-up (if applicable): Unclear time period between preoperative examination and postoperative histology. Loss-to-follow-up (if applicable): NA	Was the distribution of the (total) scores in the study sample described? (yes/no): pT-stage was not reported. *cT-stage, n:* T1: 14 T2: 13 T3: 8 T4: 4 Tumor thickness by histology, millimetres range: 1-33mm Tumor thickness by CT, millimetres range: 1-37mm Tumor thickness by MRI, millimetres range: 3.7-40 Percentage of the sample with the lowest score possible: NA	Outcome measures and effect size (include 95%CI and p-value if available): Reliability *Measurement error* The mean difference and the 95% limits of agreement could be calculated from the presented data. Bland-Altman parameters calculated from presented data for CT n=19 (CT-histopathology), mm: Mean difference: 5.93. 95% upper limit: 17.53 95% lower limit: -5.66 Bland-Altman parameters calculated from presented data for MRI n=13 (MRI-histopathology), mm: Mean difference: 8.55. 95% upper limit: 23.05 95% lower limit: -5.94	Ultrasound data was already included in the systematic review by Klein Nulent 2018 and therefore not extracted in this evidence table. T-staging according to UICC 1997 Tumors <5mm were difficult to assess by CT and MRI, according to the authors. 19/38 not detected by CT, 13/24 not detected by MRI. No agreement parameters were reported, however the mean difference and 95%LoA could be calculated from the presented data. TNM edition: TNM5 (UICC TNM5, 1997)

					Percentage of the sample with the highest score possible: NA Minimally important change/difference determined or referred? (yes/no) NA		
Verma 2019	Instruments assessed: MRI versus histopathology for **tumor thickness** (categorized, T-stage) Setting and country: Hospital, India Sampling method: Prospective over the course of 1.5 years Funding and conflicts of interest: authors declare that there was no funding and CoI	Inclusion criteria: Biopsy proven SCC of the tongue planned for surgery. Exclusion criteria: Previous history of head and neck cancer, prior surgery or radiotherapy to the neck N total at baseline: N=50 Sample characteristics[1]: *Mean age ± SD: 49 (SD not reported)* *Sex (male/female): G: 38M/12F*	Describe the assessed measurement properties and their procedures: Reliability *Measurement error* MRI was performed with 4mm slices (3T GE Signa). T1W1 axial / coronal, T2WI axial / coronal / saggital, coronal STIR, and postcontrast axial T1W were performed. Tumor thickness on MRI was measured; A reference line was drawn as the longest tumor diameter anteroposteriorly (axial view), mediolaterally (coronal view), and superoinferiorly (sagittal view). Tumor thicknes was the distance from the reference line to the deepest infiltration point	Incomplete outcome data: From sample (and subgroups if applicable) G: % SG: % Reason: How were missing data handled? NA Length of follow-up (if applicable): Time between preoperative MRI and histopathological assessment was unclear. Loss-to-follow-up (if applicable): From sample (and subgroups if applicable)	Was the distribution of the (total) scores in the study sample described? (yes/no): MRI T-stage, n: T1: 21 T2: 19 T3: 10 Histopathological T-stage, n: T1: 24 T2: 18 T3: 8 Percentage of the sample with the lowest score possible: Not relevant Percentage of the sample with the highest score possible: Not relevant Minimally important change/difference	Outcome measures and effect size (include 95%CI and p-value if available): Reliability *Measurement error* Kappa for the agreement of tumor thickness (T-stage) as measured by MRI and histopathology was not reported. However, the 3x3 table was reported from which a Kappa could be calculated: Kappa = 0.65	TNM edition: TNM7 and TNM8 (AJCC7/8), however Kappa was calculated based on TNM7 cassification

		No other characteristics provided.	and to the most projecting point of the tumor. Tumor thickness was assessed with histopathology in three dimensions. No further procedures were reported.	NA	determined or referred? (yes/no) Not relevant		
Vidiri 2019	Instruments assessed: MRI versus histopathology for **depth of invasion** and **tumor thickness** (categorized, t-stage) Setting and country: Retrospective database between 2013 and 2018 Sampling method: Database Funding and conflicts of interest: the authors declared that there were no conflicts of interest and that there was no funding received for the research, authorship, and/or publication.	Inclusion criteria: Oral tongue SCC, preoperative MRI within 3-4 weeks before surgery, availability of the depth of invasion data on the histopathological repost, presence of a measurable tumor on MRI. Exclusion criteria: Preoperative chemoradiation, mandible infiltration (T4a tumors). N total at baseline: N=43 Sample characteristics[1]: *Median age (range):*	Describe the assessed measurement properties and their procedures: Reliability MRI was performed with a 1.5T scanner (GE Optima MR 450W). T2W coronal, FSE T2W axial, pre-contrast T1W axial images were made. DWI were obtained with single-shot spin-echo and echo-planar imaging. Furthermore, post-contras T1 on de axial plane ande T1W images with liver acquisition on axial and coronal planes were made. Depth of invasion for MRI was measured using a reference line (connectin thejunctions of tumor surface and of the normal mucosa	Incomplete outcome data: From sample (and subgroups if applicable) NA How were missing data handled? NA Length of follow-up (if applicable): Time between MRI and histopathology is unclear; however, MRI was performed 3-4 weeks before surgical resection. Loss-to-follow-up (if applicable): From sample (and subgroups if applicable) NA	Was the distribution of the (total) scores in the study sample described? (yes/no): Depth of invasion for T1 tumors, median mm: Histopathology: 4 Radiologist A: 4.95 Radiologist B: 4.95 Depth of invasion for T2 tumors, median mm: Histopathology: 7 Radiologist A: 6.7 Radiologist B: 7 Depth of invasion for T3 tumors, median mm: Histopathology: 15 Radiologist A: 14.7 Radiologist B: 16 pT-stage, n: pT1: 10	Outcome measures and effect size (include 95%CI and p-value if available): Reliability *Reliability* ICC for experienced MR radiologist and histopathologic result, depth of invasion: ICC=0.90 (95%CI: 0.81-0.94) ICC for inexperienced MR radiologist and histopathologic result, depth of invasion: ICC=0.87 (95%CI: 0.76-0.92) *Measurement error* **For depth of invasion** Bland-Altman plot (MRI experienced	ICC (2,1) was used for absolute agreement. TNM edition: TNM 8 (AJCC8)

		65 (31-82) *Sex (male/female):* *G: 18M/25F* *pT-stage, n:* pT1: 10 pT2: 12 pT3: 21	surface), ignoring exophytic portions of the tumor. Invasion was meseared by drawing a line from the reference line to the deepest invasion. Radiologist assessed the cT-stage as well. MR images were assessed independently by an experienced and an inexperienced radiologist. They were blinded from histopathological results. Resected tissue samples were fixed in formalin. Embedding, sectioning and staining (hematoxylin en eosin) was done for histopathological analyses. Depth of invasion was measured by drawing a plumb line from the level of the basement membrane of the closest normal mucosa to the deepest point of invasion.		pT2: 12 pT3: 21 cT-stage experiences radiologist, n: cT1: 7 cT2: 15 cT3: 21 cT-stage inexperienced radiologist, n: cT1: 6 cT2: 19 cT3: 18 Percentage of the sample with the lowest score possible: Not relevant Percentage of the sample with the highest score possible: Not relevant Minimally important change/difference determined or referred? (yes/no) Not relevant	radiologist-histopathology), mm: Mean difference: -0.3 95% upper limit: 4.9 95% lower limit: -5.5 Bland-Altman plot (MRI inexperienced radiologist-histopathology), mm: Mean difference: -0.4 95% upper limit: 5.8 95% lower limit: -6.6 **For tumor thickness (categorized, T-stage)** Kappa for the agreement of T-stage (MRI experienced radiologist versus pathological data): K=0.74 (95%CI: 0.56-0.92) Kappa for the agreement of T-stage (MRI inexperienced radiologist versus pathological data): K=0.60 (95%CI: 0.40-0.80)	

[1] Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., … de Vet, H. C. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation, 19(4), 539-49.

Table of quality assessment for systematic reviews of RCTs and observational studies

Based on AMSTAR checklist (Shea, 2007; BMC Methodol 7: 10; doi:10.1186/1471-2288-7-10) and PRISMA checklist (Moher, 2009; PLoS Med 6: e1000097; doi:10.1371/journal.pmed1000097)

Study **First author, year**	**Appropriate and clearly focused question?[1]** **Yes/no/unclear**	**Comprehensive and systematic literature search?[2]** **Yes/no/unclear**	**Description of included and excluded studies?[3]** **Yes/no/unclear**	**Description of relevant characteristics of included studies?[4]** **Yes/no/unclear**	**Appropriate adjustment for potential confounders in observational studies?[5]** **Yes/no/unclear/ Not applicable**	**Assessment of scientific quality of included studies?[6]** **Yes/no/unclear**	**Enough similarities between studies to make combining them reasonable?[7]** **Yes/no/unclear**	**Potential risk of publication bias taken into account?[8]** **Yes/no/unclear**	**Potential conflicts of interest reported?[9]** **Yes/no/unclear**
Klein Nulent 2018	Yes Reason: Aim could have been more specified, however the inclusion/exclusion criteria are specific and probably reproducible.	Yes Reason: Multiple databases were searched. Search period was described. Syntax is available in the supplementary materials.	No (partially) Reason: Excluded studies were not described or referenced, however reasons for exclusion were provided in the study selection flow diagram.	Yes Reason: table 2 provides the study characteristics.	Not applicable	Yes Reason: QUADAS-2 tool was used.	Not applicable Reason: meta-analyses was not performed	No Reason: publication bias was not assessed. (publication bias is difficult to assess for clinimetric questions)	No Reason: It was reported for the systematic review authors, but not for the included studies.

1. **Research question (PICO) and inclusion criteria should be appropriate and predefined.**
2. **Search period and strategy should be described; at least Medline searched; for pharmacological questions at least Medline + EMBASE searched.**
3. **Potentially relevant studies that are excluded at final selection (after reading the full text) should be referenced with reasons.**
4. **Characteristics of individual studies relevant to research question (PICO), including potential confounders, should be reported.**
5. **Results should be adequately controlled for potential confounders by multivariate analysis (not applicable for RCTs).**
6. **Quality of individual studies should be assessed using a quality scoring tool or checklist (Jadad score, Newcastle-Ottawa scale, risk of bias table et cetera).**
7. **Clinical and statistical heterogeneity should be assessed; clinical: enough similarities in patient characteristics, intervention and definition of outcome measure to allow pooling? For pooled data: assessment of statistical heterogeneity using appropriate statistical tests (for example Chi-square, I^2)?**
8. **An assessment of publication bias should include a combination of graphical aids (for example funnel plot, other available tests) and/or statistical tests (for example Egger regression test, Hedges-Olken). Note: If no test values or funnel plot included, score "no". Score "yes" if mentions that publication bias could not be assessed because there were fewer than 10 included studies.**
9. **Sources of support (including commercial co-authorship) should be reported in both the systematic review and the included studies. Note: To get a "yes," source of funding or support must be indicated for the systematic review AND for each of the included studies.**

COSMIN risk of bias assessment of included studies

Reliability					
Author: Alsaffar 2019					
Instrument: clinical examination					
	Very Good	**Adequate**	**Doubtful**	**Inadequate**	**NA**
Were patients stable in the interim period on the construct to be measured?	Evidence provided that patients were stable	Assumable that patients were stable	**Unclear if patients were stable**	Patients were NOT stable	
Was the time interval appropriate?	Time interval appropriate		**Doubtful whether time interval was appropriate or time interval was not stated**	Time interval NOT appropriate	
Were the test conditions similar for the measurements? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar	
For continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	ICC calculated and model or formula of the ICC is described	ICC calculated but model or formula of the ICC not described or not optimal. Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic change has occurred	Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or WITH evidence that systematic change has occurred	No ICC or Pearson or Spearman correlations calculated	**Not applicable**
For dichotomous/nominal/ordinal scores: Was kappa calculated?	**Kappa calculated**			No kappa calculated	Not applicable
For ordinal scores: Was a weighted kappa calculated?	Weighted Kappa calculated		Unweighted Kappa calculated or not described		Not applicable
For ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	Weighting scheme described	Weighting scheme NOT described			Not applicable

Were there any other important flaws in the design or statistical methods of the study?	**No other important methodological flaws**		Other minor methodological flaws	Other important methodological flaws	

Reliability					
Author: Alsaffar 2019					
Instrument: MRI					
	Very Good	**Adequate**	**Doubtful**	**Inadequate**	**NA**
Were patients stable in the interim period on the construct to be measured?	Evidence provided that patients were stable	Assumable that patients were stable	**Unclear if patients were stable**	Patients were NOT stable	
Was the time interval appropriate?	Time interval appropriate		**Doubtful whether time interval was appropriate or time interval was not stated**	Time interval NOT appropriate	
Were the test conditions similar for the measurements? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar	
For continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	ICC calculated and model or formula of the ICC is described	ICC calculated but model or formula of the ICC not described or not optimal. Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic change has occurred	Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or WITH evidence that systematic change has occurred	No ICC or Pearson or Spearman correlations calculated	**Not applicable**
For dichotomous/nominal/ordinal scores: Was kappa calculated?	**Kappa calculated**			No kappa calculated	Not applicable
For ordinal scores: Was a weighted kappa calculated?	Weighted Kappa calculated		Unweighted Kappa calculated or not		Not applicable

			described		
For ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	Weighting scheme described	Weighting scheme NOT described			Not applicable
Were there any other important flaws in the design or statistical methods of the study?	**No other important methodological flaws**		Other minor methodological flaws	Other important methodological flaws	

Measurement error					
Author: Brouwer de Koning 2019					
Instrument: ultrasound					
	Very Good	**Adequate**	**Doubtful**	**Inadequate**	**NA**
Were patients stable in the interim period on the construct to be measured?	Patients were stable (evidence provided)	Assumable that patients were stable	**Unclear if patients were stable**	Patients were NOT stable	
Was the time interval appropriate?	Time interval appropriate		**Doubtful whether time interval was appropriate or time interval was not stated**	Time interval NOT appropriate	
Were the test conditions similar for the measurements? (e.g. type of administration, environment, instructions)	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar	
For continuous scores: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated?	**SEM, SDC, or LoA calculated**	Possible to calculate LoA from the data presented		SEM calculated based on Cronbach's alpha, or on SD from another population	Not applicable
For dichotomous/nominal/ordinal scores: Was the percentage (positive and negative) agreement calculated?	% positive and negative agreement calculated	% agreement calculated		% agreement not calculated	**Not applicable**
Were there any other important flaws in the design or statistical methods of the study?	**No other important methodological flaws**		Other minor methodological flaws	Other important methodological flaws	

Measurement error					
Author: Brouwer de Koning 2019					
Instrument: MRI					
	Very Good	**Adequate**	**Doubtful**	**Inadequate**	**NA**
Were patients stable in the interim period on the construct to be measured?	Patients were stable (evidence provided)	Assumable that	**Unclear if patients were stable**	Patients were NOT stable	

		patients were stable			
Was the time interval appropriate?	Time interval appropriate		**Doubtful whether time interval was appropriate or time interval was not stated**	Time interval NOT appropriate	
Were the test conditions similar for the measurements? (e.g. type of administration, environment, instructions)	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar	
For continuous scores: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated?	**SEM, SDC, or LoA calculated**	Possible to calculate LoA from the data presented		SEM calculated based on Cronbach's alpha, or on SD from another population	Not applicable
For dichotomous/nominal/ordinal scores: Was the percentage (positive and negative) agreement calculated?	% positive and negative agreement calculated	% agreement calculated		% agreement not calculated	**Not applicable**
Were there any other important flaws in the design or statistical methods of the study?	**No other important methodological flaws**		Other minor methodological flaws	Other important methodological flaws	

Reliability					
Author: Choi 2017					
Instrument: clinical examination (composite of endoscopic + palpation + CT or MRI)					
	Very Good	**Adequate**	**Doubtful**	**Inadequate**	**NA**
Were patients stable in the interim period on the construct to be measured?	Evidence provided that patients were stable	Assumable that patients were stable	**Unclear if patients were stable**	Patients were NOT stable	
Was the time interval appropriate?	Time interval appropriate		**Doubtful whether time interval was appropriate or time interval was not stated**	Time interval NOT appropriate	
Were the test conditions similar for the measurements? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar	

For continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	ICC calculated and model or formula of the ICC is described	ICC calculated but model or formula of the ICC not described or not optimal. Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic change has occurred	Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or WITH evidence that systematic change has occurred	No ICC or Pearson or Spearman correlations calculated	**Not applicable**
For dichotomous/nominal/ordinal scores: Was kappa calculated?	**Kappa calculated**			No kappa calculated	Not applicable
For ordinal scores: Was a weighted kappa calculated?	Weighted Kappa calculated		Unweighted Kappa calculated or not described		Not applicable
For ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	Weighting scheme described	Weighting scheme NOT described			Not applicable
Were there any other important flaws in the design or statistical methods of the study?	**No other important methodological flaws**		Other minor methodological flaws	Other important methodological flaws	

Reliability					
Author: Goel 2016					
Instrument: clinical examination					
	Very Good	**Adequate**	**Doubtful**	**Inadequate**	**NA**
Were patients stable in the interim period on the construct to be measured?	Evidence provided that patients were stable	Assumable that patients were stable	**Unclear if patients were stable**	Patients were NOT stable	
Was the time interval appropriate?	Time interval appropriate		**Doubtful whether time interval was appropriate or time interval was not stated**	Time interval NOT appropriate	

Were the test conditions similar for the measurements? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar	
For continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	ICC calculated and model or formula of the ICC is described	ICC calculated but model or formula of the ICC not described or not optimal. Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic change has occurred	Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or WITH evidence that systematic change has occurred	No ICC or Pearson or Spearman correlations calculated	**Not applicable**
For dichotomous/nominal/ordinal scores: Was kappa calculated?	**Kappa calculated**			No kappa calculated	Not applicable
For ordinal scores: Was a weighted kappa calculated?	Weighted Kappa calculated		Unweighted Kappa calculated or not described		Not applicable
For ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	Weighting scheme described	Weighting scheme NOT described			Not applicable
Were there any other important flaws in the design or statistical methods of the study?	**No other important methodological flaws**		Other minor methodological flaws	Other important methodological flaws	

Reliability					
Author: Goel 2016					
Instrument: MRI					
	Very Good	**Adequate**	**Doubtful**	**Inadequate**	**NA**
Were patients stable in the interim period on the construct to be measured?	Evidence provided that patients were stable	Assumable that patients were stable	**Unclear if patients were stable**	Patients were NOT stable	
Was the time interval appropriate?	Time interval appropriate		**Doubtful whether time interval was**	Time interval NOT appropriate	

			appropriate or time interval was not stated		
Were the test conditions similar for the measurements? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar	
For continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	ICC calculated and model or formula of the ICC is described	ICC calculated but model or formula of the ICC not described or not optimal. Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic change has occurred	Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or WITH evidence that systematic change has occurred	No ICC or Pearson or Spearman correlations calculated	**Not applicable**
For dichotomous/nominal/ordinal scores: Was kappa calculated?	**Kappa calculated**			No kappa calculated	Not applicable
For ordinal scores: Was a weighted kappa calculated?	Weighted Kappa calculated		Unweighted Kappa calculated or not described		Not applicable
For ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	Weighting scheme described	Weighting scheme NOT described			Not applicable
Were there any other important flaws in the design or statistical methods of the study?	**No other important methodological flaws**		Other minor methodological flaws	Other important methodological flaws	

Reliability					
Author: Iida 2018					
Instrument: ultrasound					
	Very Good	**Adequate**	**Doubtful**	**Inadequate**	**NA**
Were patients stable in the interim period on the construct to be measured?	Evidence provided that patients were	Assumable that	**Unclear if patients were stable**	Patients were NOT stable	

	stable	patients were stable			
Was the time interval appropriate?	Time interval appropriate		**Doubtful whether time interval was appropriate or time interval was not stated**	Time interval NOT appropriate	
Were the test conditions similar for the measurements? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar	
For continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	ICC calculated and model or formula of the ICC is described	ICC calculated but model or formula of the ICC not described or not optimal. Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic change has occurred	Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or WITH evidence that systematic change has occurred	No ICC or Pearson or Spearman correlations calculated	**Not applicable**
For dichotomous/nominal/ordinal scores: Was kappa calculated?	**Kappa calculated**			No kappa calculated	Not applicable
For ordinal scores: Was a weighted kappa calculated?	Weighted Kappa calculated		Unweighted Kappa calculated or not described		Not applicable
For ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	Weighting scheme described	Weighting scheme NOT described			Not applicable
Were there any other important flaws in the design or statistical methods of the study?	**No other important methodological flaws**		Other minor methodological flaws	Other important methodological flaws	

Measurement error					
Author: Mao 2019					
Instrument: MRI					
	Very Good	**Adequate**	**Doubtful**	**Inadequate**	**NA**
Were patients stable in the interim period on the construct to be measured?	Patients were stable (evidence provided)	**Assumable that patients were stable**	Unclear if patients were stable	Patients were NOT stable	
Was the time interval appropriate?	**Time interval appropriate**		Doubtful whether time interval was appropriate or time interval was not stated	Time interval NOT appropriate	
Were the test conditions similar for the measurements? (e.g. type of administration, environment, instructions)	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar	
For continuous scores: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated?	**SEM, SDC, or LoA calculated**	Possible to calculate LoA from the data presented		SEM calculated based on Cronbach's alpha, or on SD from another population	Not applicable
For dichotomous/nominal/ordinal scores: Was the percentage (positive and negative) agreement calculated?	% positive and negative agreement calculated	% agreement calculated		% agreement not calculated	**Not applicable**
Were there any other important flaws in the design or statistical methods of the study?	**No other important methodological flaws**		Other minor methodological flaws	Other important methodological flaws	

Measurement error					
Author: Shintani 2001					
Instrument: CT					
	Very Good	**Adequate**	**Doubtful**	**Inadequate**	**NA**
Were patients stable in the interim period on the construct to be measured?	Patients were stable (evidence provided)	Assumable that patients were stable	**Unclear if patients were stable**	Patients were NOT stable	
Was the time interval appropriate?	Time interval appropriate		**Doubtful whether time interval was appropriate or time interval was not stated**	Time interval NOT appropriate	

Were the test conditions similar for the measurements? (e.g. type of administration, environment, instructions)	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar	
For continuous scores: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated?	**SEM, SDC, or LoA calculated**	Possible to calculate LoA from the data presented		SEM calculated based on Cronbach's alpha, or on SD from another population	Not applicable
For dichotomous/nominal/ordinal scores: Was the percentage (positive and negative) agreement calculated?	% positive and negative agreement calculated	% agreement calculated		% agreement not calculated	**Not applicable**
Were there any other important flaws in the design or statistical methods of the study?	**No other important methodological flaws**		Other minor methodological flaws	Other important methodological flaws	

Measurement error					
Author: Shintani 2001					
Instrument: MRI					
	Very Good	**Adequate**	**Doubtful**	**Inadequate**	**NA**
Were patients stable in the interim period on the construct to be measured?	Patients were stable (evidence provided)	Assumable that patients were stable	**Unclear if patients were stable**	Patients were NOT stable	
Was the time interval appropriate?	Time interval appropriate		**Doubtful whether time interval was appropriate or time interval was not stated**	Time interval NOT appropriate	
Were the test conditions similar for the measurements? (e.g. type of administration, environment, instructions)	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar	
For continuous scores: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated?	**SEM, SDC, or LoA calculated**	Possible to calculate LoA from the data presented		SEM calculated based on Cronbach's alpha, or on SD from another population	Not applicable
For dichotomous/nominal/ordinal scores: Was the percentage (positive and negative) agreement calculated?	% positive and negative agreement calculated	% agreement calculated		% agreement not calculated	**Not applicable**

Were there any other important flaws in the design or statistical methods of the study?	**No other important methodological flaws**		Other minor methodological flaws	Other important methodological flaws	

Measurement error					
Author: Nair 2018					
Instrument: Ultrasound					
	Very Good	**Adequate**	**Doubtful**	**Inadequate**	**NA**
Were patients stable in the interim period on the construct to be measured?	Patients were stable (evidence provided)	Assumable that patients were stable	**Unclear if patients were stable**	Patients were NOT stable	
Was the time interval appropriate?	Time interval appropriate		**Doubtful whether time interval was appropriate or time interval was not stated**	Time interval NOT appropriate	
Were the test conditions similar for the measurements? (e.g. type of administration, environment, instructions)	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar	
For continuous scores: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated?	**SEM, SDC, or LoA calculated**	Possible to calculate LoA from the data presented		SEM calculated based on Cronbach's alpha, or on SD from another population	Not applicable
For dichotomous/nominal/ordinal scores: Was the percentage (positive and negative) agreement calculated?	% positive and negative agreement calculated	% agreement calculated		% agreement not calculated	**Not applicable**
Were there any other important flaws in the design or statistical methods of the study?	**No other important methodological flaws**		Other minor methodological flaws	Other important methodological flaws	

Reliability					
Author: Verma 2019					
Instrument: MRI					
	Very Good	**Adequate**	**Doubtful**	**Inadequate**	**NA**
Were patients stable in the interim period on the construct to be measured?	Evidence provided that patients were stable	Assumable that patients were stable	**Unclear if patients were stable**	Patients were NOT stable	
Was the time interval appropriate?	Time interval appropriate		**Doubtful whether time interval was**	Time interval NOT appropriate	

			appropriate or time interval was not stated		
Were the test conditions similar for the measurements? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar	
For continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	ICC calculated and model or formula of the ICC is described	ICC calculated but model or formula of the ICC not described or not optimal. Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic change has occurred	Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or WITH evidence that systematic change has occurred	No ICC or Pearson or Spearman correlations calculated	**Not applicable**
For dichotomous/nominal/ordinal scores: Was kappa calculated?	**Kappa calculated**			No kappa calculated	Not applicable
For ordinal scores: Was a weighted kappa calculated?	Weighted Kappa calculated		Unweighted Kappa calculated or not described		Not applicable
For ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	Weighting scheme described	Weighting scheme NOT described			Not applicable
Were there any other important flaws in the design or statistical methods of the study?	**No other important methodological flaws**		Other minor methodological flaws	Other important methodological flaws	

Measurement error					
Author: Vidiri 2019					
Instrument: MRI (DOI)					
	Very Good	**Adequate**	**Doubtful**	**Inadequate**	**NA**
Were patients stable in the interim period on the construct to be measured?	Patients were stable (evidence provided)	**Assumable that**	Unclear if patients were stable	Patients were NOT stable	

		patients were stable			
Was the time interval appropriate?	**Time interval appropriate**		Doubtful whether time interval was appropriate or time interval was not stated	Time interval NOT appropriate	
Were the test conditions similar for the measurements? (e.g. type of administration, environment, instructions)	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar	
For continuous scores: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated?	**SEM, SDC, or LoA calculated**	Possible to calculate LoA from the data presented		SEM calculated based on Cronbach's alpha, or on SD from another population	Not applicable
For dichotomous/nominal/ordinal scores: Was the percentage (positive and negative) agreement calculated?	% positive and negative agreement calculated	% agreement calculated		% agreement not calculated	**Not applicable**
Were there any other important flaws in the design or statistical methods of the study?	**No other important methodological flaws**		Other minor methodological flaws	Other important methodological flaws	

Table of excluded studies

Authors and year	Reason for exclusion
Angelelli 2017	No parameters of interest reported
Bashir 2011	No parameters of interest reported
Chammas 2011	Included in the systematic review by Klein Nulent 2018
Iwai 2002	No parameters of interest reported
Jayasankaran2017	No parameters of interest reported
Jung 2009	No parameters of interest reported
Junn 2017	No parameters of interest reported
Kodama 2010	Included in the systematic review by Klein Nulent 2018
Koopaie 2014	Article not available in English
Lam 2004	No parameters of interest reported
Lwin 2012	No parameters of interest reported
Madana 2015	No parameters of interest reported
Morand 2019	No parameters of interest reported
Moreno 2017	No parameters of interest reported
Park 2011	No parameters of interest reported
Preda 2006	No parameters of interest reported
Sarode 2010	Letter to the editor
Sarode 2012	Educational/opinion paper
Shintni 1997	No parameters of interest reported
Songra 2006	Included in the systematic review by Klein Nulent 2018
Weimar 2018	No parameters of interest reported
Yesuratnam 2014	Included in the systematic review by Klein Nulent 2018

Yuen 2008	No parameters of interest reported
Zhou 2008	Article not available in English

Literature search strategy

Uitgangsvraag: Richtlijn HHT. Doe dient de invasiediepte van de tumor (mondholtecarcinoom) (preoperatief) bepaald te worden	
Database(s): Ovid/Medline, Embase	Datum: 12-11-2019
Periode: niet van toepassing	Talen: nvt
Toelichting: Alle sleutelartikelen zijn gevonden, zie set 9 tot en met 16.	

Studietype	Inclusief dubbele referenties	Ontdubbeld
SR	12	12
RCT		
Diagnostisch	301	299
Totaal		**311**

Zoekverantwoording

Ovid/Medline

1 exp Mouth Neoplasms/ (67203)
2 ((mouth or "mouth cavity" or oral or intraoral or "buccal mucosa" or jaw or palat* or mandibular or "mammary analogue secretory" or tongue or gingiva* or tonsil* of pharynx or hypopharynx or nasopharynx or oropharynx or "parotid gland*or salivary duct*" or "salivary gland*" or lympoepithelioma) adj2 (tumor* or carcinoma* or tumour* or squamous or cancer* or neoplasm*)).ti,ab,kf. (40652)
3 (exp Neoplasm Invasiveness/ or (invasive* or invasion).ti,ab,kf.) adj3 (tumor* or carcinoma* or tumour* or squamous or cancer* or neoplasm* or "buccal mucosaor tongue" or tonsil* or pharynx of hypopharynx or naospharynx or "oropharynx of parotid glandor salivary ductor salivary gland*" or depth).ti,ab,kf. (38843)
4 1 or 2 (81357)
5 3 and 4 (907)
6 exp "Sensitivity and Specificity"/ or (specificit* of screening or accura* or reference value* or false positive of false negative or predictive value or roc or likelyhood* or likelihood*).ti,ab,kf. (1391046)
7 5 and 6 (81)
8 (meta-analysis/ or meta-analysis as topic/ or (meta adj analy$).tw. or ((systematic* or literature) adj2 review$1).tw. or (systematic adj overview$1).tw. or exp "Review Literature as Topic"/ or cochrane.ab. or cochrane.jw. or embase.ab. or medline.ab. or (psychlit or psyclit).ab. or (cinahl or cinhal).ab. or cancerlit.ab. or ((selection criteria or data extraction).ab. and "review"/)) not (Comment/ or Editorial/ or Letter/ or (animals/ not humans/)) (419900)
9 7 and 8 (5)
10 7 not 9 (76)

Embase

No.	Query	Results
#19	#17 NOT #18	**234**
#18	#6 AND #17	**7**
#17	#5 NOT 'conference abstract':it	**241**
#16	#5 AND #15	**1**
#15	klein AND intraoral AND ultrasonography AND thickness AND cancer AND 2017	**1**

#14	#5 AND #13	**1**
#13	lwin AND accuracy AND mri AND tumour AND thickness AND nodal AND stage AND oral AND 2012	**1**
#12	alsaffar AND correlation AND clinical AND mri AND invasion AND oral AND tongue AND 2016	**1**
#11	#5 AND #10	**1**
#10	park AND diagnostic AND accuracy AND tumor AND invasion AND imaging AND 2011 AND can cer	**19**
#9	#5 AND #8	**1**
#8	vidiri AND depth AND of AND invasion AND in AND staging AND 2019	**1**
#7	#5 AND #6	**7**
#6	('meta analysis'/de OR cochrane:ab OR embase:ab OR psycinfo:ab OR cinahl:ab OR medline:ab OR ((systematic NEAR/1 (review OR overview)):ab,ti) OR ((meta NEAR/1 analy*):ab,ti) OR metaanalys*:ab,ti OR 'data extraction':ab OR cochrane:jt OR 'systematic review'/de) NOT (('animal experiment'/exp OR 'animal model'/exp OR 'nonhuman'/exp) NOT 'human'/exp)	**461633**
#5	#1 AND #2 AND #3 AND #4	**261**
#4	'sensitivity and specificity'/exp OR 'screening'/exp OR 'reference value'/exp OR 'diagnostic error'/exp OR 'predictive value'/exp OR 'receiver operating characteristic'/exp OR specificit*:ab,ti OR screening:ab,ti OR accura*:ab,ti OR 'reference value*':ab,ti OR 'false positive':ab,ti OR 'false negative':ab,ti OR 'predictive value*':ab,ti OR roc:ab,ti OR likelyhood*:ab,ti OR likelihood*:ab,ti	**2954561**
#3	('echography'/exp OR 'doptone':ti,ab OR 'duplex echography':ti,ab OR 'echogram':ti,ab OR 'echography':ti,ab OR 'echoscopy':ti,ab OR 'echosound':ti,ab OR 'high resolution echography':ti,ab OR 'scanning, ultrasonic':ti,ab OR 'sonogram':ti,ab OR 'sonography':ti,ab OR 'ultrasonic detection':ti,ab OR 'ultrasonic diagnosis':ti,ab OR 'ultrasonic echo':ti,ab OR 'ultrasonic examination':ti,ab OR 'ultrasonic scanning':ti,ab OR 'ultrasonic scintillation':ti,ab OR 'ultrasonography':ti,ab OR 'ultrasound diagnosis':ti,ab OR 'ultrasound scanning':ti,ab) AND intraoral:ti,ab OR 'palpation'/exp OR 'palpation':ti,ab OR 'diagnostic imaging'/exp OR 'diagnostic imaging':ti,ab OR 'imaging, diagnostic':ti,ab OR 'imaging, medical':ti,ab OR 'medical imaging':ti,ab OR 'ophthalmo diaphanoscopy':ti,ab OR 'nuclear magnetic resonance imaging'/exp OR 'expert plus':ti,ab OR 'expert plus 1t':ti,ab OR 'mri':ti,ab OR 'nmr imaging':ti,ab OR 'imaging, magnetization transfer':ti,ab OR 'magnetic resonance imaging':ti,ab OR 'magnetic resonance tomography':ti,ab OR 'magnetization transfer imaging':ti,ab OR 'mr imaging':ti,ab OR 'nuclear magnetic resonance imaging':ti,ab OR 'pet mri':ti,ab OR 'positron emission tomography'/exp OR 'pet scan':ti,ab OR 'pet scanning':ti,ab OR 'p.e.t.':ti,ab OR 'positron emission tomographic scan':ti,ab OR 'positron emission tomographic scanning':ti,ab OR 'positron emission tomography':ti,ab OR 'positron tomography':ti,ab OR 'positron-emission tomography':ti,ab OR 'pet ct':ti,ab OR 'computer assisted tomography'/exp OR 'cat scan':ti,ab OR 'cat scanning':ti,ab OR 'computed tomographic scan':ti,ab OR 'computed tomography':ti,ab OR 'computed tomography	**1903478**

	scan':ti,ab OR 'computer assisted tomography':ti,ab OR 'computer tomography':ti,ab OR 'computerised axial tomography':ti,ab OR 'computerised tomography':ti,ab OR 'computerized axial tomography':ti,ab OR 'computerized tomography':ti,ab OR 'computerized tomography scan':ti,ab OR 'digital examination':ti,ab	
#2	'tumor invasion'/exp OR 'cancer invasion':ti,ab OR 'invasive cancer':ti,ab OR 'neoplasm invasion':ti,ab OR 'neoplasm invasiveness':ti,ab OR 'tumor invasion':ti,ab OR 'tumour invasion':ti,ab OR (((tumor* OR tumour* OR neoplasm* OR cancer* OR carcinoma*) NEAR/3 thick*):ti,ab) OR (depth:ti AND invasion:ti) OR (depth:ab AND invasion:ab)	**131590**
#1	'mouth tumor'/exp OR 'buccal mucosa tumor':ti,ab OR 'buccal mucosa tumour':ti,ab OR 'intraoral tumor':ti,ab OR 'intraoral tumour':ti,ab OR 'mouth cavity tumor':ti,ab OR 'mouth cavity tumour':ti,ab OR 'mouth neoplasm*':ti,ab OR 'mouth tumor*':ti,ab OR 'mouth tumour*':ti,ab OR 'oral cavity tumor*':ti,ab OR 'oral cavity tumour*':ti,ab OR 'oral mucosa tumor*':ti,ab OR 'oral mucosa tumour*':ti,ab OR 'oral tumor':ti,ab OR 'oral tumour*':ti,ab OR 'oral neoplasm*':ti,ab OR 'oral carcinoma*':ti,ab OR ('oral cavity':ti AND carcinoma*:ti) OR ('oral cavity':ab AND carcinoma*:ab) OR 'mouth cancer'/exp OR 'intraoral cancer':ti,ab OR 'mouth cancer':ti,ab OR 'mouth mucosa cancer':ti,ab OR 'oral cancer':ti,ab OR 'oral cavity cancer':ti,ab OR ((oral:ti OR mouth:ti) AND squamous:ti) OR ((oral:ab OR mouth:ab) AND squamous:ab) OR 'oral cavity squamous cell carcinoma'/exp	**132423**

Module 1.2 Diagnostiek botinvasie mandibula

Uitgangsvraag

Hoe dient botinvasie van de mandibula preoperatief bepaald te worden?

Inleiding

Mandibulaire botinvasie door een mondholtecarcinoom is geassocieerd met een slechtere lokale controle. Botinvasie zou daarom behandeld dienen te worden met een chirurgische resectie van een deel van de mandibula, wat meer functieverlies voor de patiënt kan veroorzaken. Vanwege de verstrekkende consequenties van een chirurgische resectie is het belangrijk om met beeldvorming vóór de ingreep betrouwbaar te kunnen bepalen of er sprake is van mandibulaire botinvasie.

Search and select

A systematic review of the literature was performed to answer the following question:

> What is the diagnostic accuracy of an orthopantogram (OPG), Cone Beam Computed Tomography (CBCT), Computed Tomography (CT), Positron Emission Tomography-Computed Tomography (PET-CT), Single-Photon Emission Computed Tomography (SPECT), Dual Energy Computed Tomography (DECT), Magnetic Resonance Imaging (MRI), or diagnostic algorithms with a pathological assessment as reference to diagnose mandibular bone invasion of a tumor preoperatively in patients with oral cancer?

P: patients with an oral cavity carcinoma at risk for mandibular bone invasion;
I: preoperative diagnosis of mandibular invasion with OPG, CBCT, CT, PET-CT, SPECT, DECT, MRI, or with a diagnostic algorithm;
C: comparison of modalities;
R: postoperative pathological assessment;
O: sensitivity, specificity, positive predictive value, negative predictive value.

A priori, the working group did not define cortical/medullary bone invasion but used the definitions used in the studies.

<u>Relevant outcome measures</u>

The guideline development group considered sensitivity and negative predictive value as a critical outcome measure for decision making; and specificity and positive predictive value as an important outcome measure for decision making.

<u>Search and select (Methods)</u>

The databases Medline (via OVID) and Embase (via embase.com) were searched with relevant terms for primary diagnostic studies of diagnostic algorithms until 18th of November 2019. The systematic literature search resulted in 56 hits. Studies reporting diagnostic algorithms were selected on the following criteria: the population were patients with oral cavity carcinomas, the patients were suspected of mandibular bone invasion, the mandibular invasion was preoperatively diagnosed by using a diagnostic algorithm, the diagnostic accuracy was reported, the reference standard was a postoperative pathological assessment. Studies reporting an algorithm were excluded when the (imaging) data were collected before the year 2000. Based on title and abstract, nineteen studies were initially selected. Eighteen studies were excluded after reading the full-text (see the table with reasons for exclusion under the tab Evidence tables). One study reporting diagnostic algorithms was included.

A second search was performed. The databases Medline (via OVID) and Embase (via embase.com) were searched with relevant terms for diagnostic test accuracy systematic reviews of imaging modalities until 20th of November 2019. The systematic literature search resulted in 22 hits. Studies reporting the diagnostic test accuracy of imaging modalities were selected based on the following criteria: the population were patients with oral cavity carcinomas, the patients were suspected of mandibular bone invasion, the mandibular invasion was preoperatively diagnosed with OPG/CBCT/CT/PET-CT/SPECT/DECT/MRI, the diagnostic accuracy was reported, the reference standard was a postoperative pathological assessment. Based on title and abstract, eleven systematic reviews were initially selected. Ten systematic reviews were excluded after reading the full-text (see the table with reasons for exclusion under the tab Evidence tables). One systematic review reporting the diagnostic accuracy of imaging modalities was included.

The details of both search strategies are depicted under the tab Methods.

Results

Two studies were included in the analysis of the literature (one study for diagnostic algorithms and one diagnostic test accuracy systematic review). Important study characteristics and results are summarized in the evidence tables. The assessment of the risk of bias is summarized in the risk of bias tables (under the tab Evidence tables).

Summary of literature

Description of the systematic review for the diagnostic accuracy of imaging modalities

Qiao (2018) performed a diagnostic test accuracy systematic review of imaging modalities for the diagnosis of mandibular bone invasion. On the first of November 2017 the following databases were searched by Qiao (2018): MEDLINE, CINAHL, Latin American and Caribbean Health Sciences, Chinese Biomedical Literature Databases, China National Knowledge Infrastructure, VIP database, and the Wanfang Database. Grey literature was searched in Science Paper Online, System for Information on Grey Literature in Europe, and the WHO International Clinical Trials Registry Platform. Twenty-one Chinese journals were hand-searched for additional relevant studies (journals were not referenced). Studies were included when designed as a cohort, participants were diagnosed with oral or head and neck cancer with preoperative biopsy and mandibulectomy during surgery, the index test of interest was used (CT, MRI, CBCT, OPG, PET-CT, SPECT, BS, US), pathological diagnosis was used as a reference test, the target condition was mandible invasion by the tumor, and when outcomes of interest were reported (or could be calculated). Their search and selection resulted in the inclusion of 49 unique studies (of which 45 were relevant to the PICRO in this guideline module). All of the included studies recruited patients with oral cavity carcinomas, while less than half the studies also recruited some patients with tumor locations outside the oral cavity. Twenty-six studies exclusively selecting oral cavity carcinomas recruited 1372 patients in total. Twenty-three studies recruiting patients with tumors at oral cavity sites as well as head and neck sites included the following tumor locations: cheek (13 studies), lymph node (1 study), tonsil (6 studies), mandible (2 studies), submandibular trigone (1 study), submandibular glad (1 study), oropharynx (5 studies), and/or pharynx (1 study). Two authors independently assessed the risk of bias of the individual studies with the QUADAS-2 instrument. Results of the studies were pooled (statistical model was not mentioned) and a meta-regression was performed, if possible, to assess the observed heterogeneity between studies. The mean age (when reported) of patients varied from 51.5 to 73.6 years. The number of male patients (when reported) participating in studies varied from 50% to 88.2%. The prevalence of mandibular invasion in the included studies varied from 17.6% to 95.2%. The used tracers for SPECT in the 14

included studies were: ^{99m}Technetium mythelene diphosphonate (n=7), ^{99m}Technetium hydroxymethelene di phosphonate (n=4), ^{99m}Technetium dicarboxy propan (n=1), ^{99m}Technetium 3,3-disphosphono-1,2-propanedicarboxylic acid (n=1), ^{201}TI-chloride (n=1), ^{99m}Technetium-bisphosphonate (n=1, unclear which specific type), unclear (n=2). Two studies used two different tracers for separate groups in the sample (^{99m}Technetium mythelene diphosponate or ^{99m}Technetium hydroxymethylene diphosponate). One study used a dual isotope protocol (^{99m}Technetium hydroxymethylene diphosphonate and ^{201}TI-chloride).

Description of studies with diagnostic algorithms

Van Cann (2008) described eight different diagnostic algorithms using SPECT, CT, and MRI to diagnose mandibular bone invasion in patients with oral cavity carcinomas. Data from 67 patients (62.7% male) were analyzed. The mean age was 64 years (range: 43-84). Patients had tumours at the following sites: floor of mouth (n=31), retromolar area (n=20), lower alveolar process (n=13), or cheek mucosa (n=3). There were two assessors for each imaging modality. The assessors were blinded from the results of the imaging modalities they did not assess. A positive result for CT was defined as the absence of cortex adjacent to an abnormal tissue mass. A positive result for MRI was defined as the replacement of the hypointense signal of cortical bone by the signal intensity of a tumour on both the SE T2-weighted and SE T1-weighted images, or as a replacement of hyperintense signal of medullary bone by the tumour intensity signal. No criteria were provided for a positive SPECT. The reference test was a histopathological assessment. Cortical bone invasion was defined as the replacement of bone by an advancing tumour front, without invasion to cancellous spaces, the dental canal, or the periodontal ligament. Medullary invasion was defined as the diffuse growth through the cortex into cancellous bone, the dental canal, or the periodontal ligament. The absence of bone invasion was defined as a continuous periosteal layer separating the tumor from bone. Eight algorithms were presented:

- **Algorithm I:** start with SPECT
  - Negative SPECT → No invasion.
  - Positive SPECT, continue with MRI:
    - Negative MRI, continue with CT:
      - Negative CT → No invasion.
      - Positive CT → Invasion.
    - Positive MRI → Invasion.

- **Algorithm II:** start with SPECT:
  - Negative SPECT → No invasion.
  - Positive SPECT, continue with CT:
    - Negative CT, continue with MRI:
      - Negative MRI → No invasion.
      - Positive MRI →Invasion.
    - Positive CT → Invasion.

- **Algorithm III:** start with MRI:
  - Positive MRI → Invasion.
  - Negative MRI, continue with SPECT:
    - Positive SPECT, continue with CT:
      - Negative CT → No invasion.
      - Positive CT → Invasion.
    - Negative SPECT → No invasion.

- **Algorithm IV:** start with MRI:
  - Positive MRI → Invasion.
  - Negative MRI, continue with CT:
    - Negative CT, continue with SPECT:
      - Negative SPECT → No invasion.
      - Positive SPECT → Invasion.
    - Positive CT→ invasion.

- **Algorithm V:** start with CT:
  - Positive CT → Invasion.
  - Negative CT, continue with SPECT:
    - Positive SPECT, continue with MRI:
      - Negative MRI → No invasion.
      - Positive MRI → Invasion.
    - Negative SPECT → No invasion.

- **Algorithm VI:** start with CT:
  - Positive CT -> Invasion.
  - Negative CT, continue with MRI:
    - Negative MRI, continue with SPECT:
      - Negative SPECT → No invasion.
      - Positive SPECT → Invasion.
    - Positive MRI → Invasion.

- **Algorithm VII:** start with MRI:
  - Positive MRI → Invasion.
  - Negative MRI, continue with SPECT:
    - Positieve SPECT → Invasion
    - Negatieve SPECT → No invasion

- **Algorithm VIII:** start with CT:
  - Positive CT → Invasion.
  - Negative CT, continue with SPECT:
    - Positive SPECT → Invasion.
    - Negative SPECT → No invasion.

Results

Diagnostic test accuracy of imaging modalities

Orthopantogram (OPG)

Sensitivity

From 15 studies (n=772, as described in the systematic review's characteristics table) data were pooled by Qiao (2018) for the sensitivity of OPG in diagnosing mandibular bone invasion. A pooled sensitivity estimate of 0.75 (95%CI: 0.67 to 0.82, I^2=62%) was reported. For medullary invasion, Qiao (2018) included 1 study (n=29, as described in the study characteristics table) where a sensitivity of 0.63 was found (95%CI not reported).

Specificity

A pooled specificity estimate of 0.83 (95%CI: 0.79 to 0.86, I^2=19%) was calculated by Qiao (2018) for OPG in diagnosing mandibular bone invasion. Data were pooled from 15 studies (n=772, as described in the systematic review's characteristics table). Qiao (2018) included 1

study (n=29, as described in the study characteristics table) for the specificity of OPG in diagnosing medullary invasion. A specificity of 0.90 was presented (95%CI not reported).

Positive predictive value
No studies were included that reported the positive predictive value for OPG.

Negative predictive value
No studies were included that reported the negative predictive value for OPG.

Cone Beam Computed Tomography (CBCT):
Sensitivity
Data from 5 studies were pooled (n=557, as described in the systematic review's characteristics table) for CBCT diagnosing mandibular invasion (Qiao, 2018). The pooled sensitivity was 0.90 (95%CI: 0.85 to 0.93, I^2=0%).

Specificity
Qiao (2018) pooled data from 5 studies (n=557, as described in the systematic review's characteristics table) for the specificity of CBCT in diagnosing mandibular bone invasion. A pooled specificity of 0.85 (95%CI: 0.62 to 0.95, I^2=80%) was reported.

Positive predictive value
No studies were included that reported the positive predictive value for CBCT.

Negative predictive value
No studies were included that reported the negative value for CBCT.

Computed Tomography (CT)
Sensitivity
From 35 studies (n=1908, as described in the systematic review's characteristics table) data was pooled by Qiao (2018) for the sensitivity of CT diagnosing mandibular bone invasion. The pooled sensitivity was 0.73 (95%CI: 0.66 to 0.80, I^2=70%). Qiao (2018) also pooled data from 4 studies (n=145, as described in the study characteristics table) for the sensitivity of CT in diagnosing mandibular medullary invasion. A pooled sensitivity of 0.85 (95%CI: 0.43 to 0.98, I^2=83%) was found.

Specificity
Qiao (2018) calculated a pooled estimate of the specificity of CT diagnosing mandibular bone invasion from 35 studies (n=1908, as described in the systematic review's characteristics table). From these 35 studies, 9 were published before the year 2000 (publication range: 1990 and 1998, n=397 participants) and 26 were published since the year 2000 (publication range: 2000-2014, n=1511 participants). A pooled specificity of 0.91 (95%CI: 0.88 to 0.94, I^2=49%) was reported. The pooled specificity for CT diagnosing medullary invasion was calculated from 5 studies (n=145, as described in the study characteristics table) and was 0.86 (95%CI: 0.73 to 0.93, I^2=32%).

Positive predictive value
No studies were included that reported the positive predictive value for CT.

Negative predictive value
No studies were included that reported the negative predictive value for CT.

Positron Emission Tomography-Computed Tomography (PET-CT)
Sensitivity
Four studies (n=114, as described in the systematic review's characteristics table) were pooled by Qiao (2018) to calculate a summary sensitivity for diagnosing mandibular bone invasion. All of these four studies were published after the year 2000 (range: 2005 and 2011). A pooled sensitivity of 0.90 (95%CI: 0.58 to 0.85, I^2= 64%) was reported. Qiao (2018) also included 2 studies for diagnosing medullary invasion with PET-CT. The reported sensitivity in both studies were 0.78 and 1.00, respectively (95% CI's were not reported).

Specificity
Qiao (2018) pooled 4 studies (n= 114, as described in the systematic review's characteristics table) to calculate a summary specificity. The four studies were published between 2005 and 2011. A summary specificity of 0.89 (95%CI: 0.77 to 0.96) was reported for PET-CT diagnosing mandibular bone invasion. Qiao (2018) also included two studies for diagnosing medullary invasion with PET-CT. The reported specificity in both studies were 0.14 and 0.86, respectively (95% CI's were not reported).

Positive predictive value
No studies were included that reported the positive predictive value for PET-CT.

Negative predictive value
No studies were included that reported the negative predictive value for PET-CT.

Single-Photon Emission Computed Tomography (SPECT)
Sensitivity
Qiao (2018) used data from 13 studies on the diagnostic accuracy of SPECT for diagnosing mandibular bone invasion. There was data from 858 patients (as described in the systematic review's characteristics table of Qiao (2018). Data was pooled and a sensitivity of 0.97 (95%CI: 0.92 to 0.99, I^2=72%) was reported.

Specificity
Qiao (2018) reported a pooled specificity of 0.69 (95%CI: 0.52 to 0.82, I^2=79%) for SPECT diagnosing mandibular bone invasion. Data from 13 studies (n=858, as described in the systematic review's characteristics table) was pooled.

Positive predictive value
No studies were included that reported the positive predictive value for SPECT.

Negative predictive value
No studies were included that reported the negative predictive value for SPECT.

Dual Energy Computed Tomography (DECT):
No studies were included that reported the diagnostic test accuracy of DECT.

Magnetic Resonance Imaging (MRI)
Sensitivity
For the sensitivity of MRI in diagnosing mandibular bone invasion, Qiao (2018) pooled data from 18 studies (n=820, as described in the systematic review's characteristics table). A pooled sensitivity of 0.88 (95%CI: 0.78 to 0.94, I^2=76%) was presented. For diagnosing medullary invasion with MRI, data from 7 studies (n=311, as described in the study

characteristics table) were pooled by Qiao (2018). The pooled sensitivity was 0.93 (95%CI: 0.81 to 0.98, I^2=79%).

Specificity
The pooled specificity of MRI for diagnosing mandibular bone invasion was calculated by Qiao (2018) from data in 18 studies (n=820, as described in the systematic review's characteristics table). The pooled specificity was 0.90 (95%CI: 0.80 to 0.95, I^2=81%). Qiao (2018) also pooled data for the specificity of MRI in diagnosing medullary invasion. From 7 studies (n=311, as described in the study characteristics table), a specificity of 0.84 (95%CI: 0.60 to 0.95, I^2=79%) was found.

Positive predictive value
No studies were included that reported the positive predictive value for MRI.

Negative predictive value
No studies were included that reported the positive predictive value for MRI.

Algorithms for bone invasion
Sensitivity
The sensitivity of the algorithms was not reported (Van Cann, 2008). From the reported data about the classifications of the algorithms (i.e. true positives, false positives, false negatives, true negatives) a sensitivity was calculated per algorithm. The sensitivity of the algorithms varied between 0.77 and 1.00. See Table 1.5 for an overview per algorithm and Figure 1.6 for a graphical representation.

Specificity
The specificity of the algorithms was not reported (Van Cann, 2008). From the reported data about the classifications of the algorithms (i.e. true positives, false positives, false negatives, true negatives) a specificity was calculated per algorithm. The specificity of the algorithms varied between 0.57 and 1.00. See Table 1.5 for an overview per algorithm and Figure 1.6 for a graphical representation.

Positive predictive value
The positive predictive value of the algorithms was not reported (Van Cann, 2008). From the reported data about the classifications of the algorithms (i.e. true positives, false positives, false negatives, true negatives) a positive predictive value was calculated per algorithm. The positive predictive value of the algorithms varied between 0.80 and 1.00. See Table 1.5 for an overview per algorithm.

Negative predictive value
The negative predictive value of the algorithms was not reported (Van Cann, 2008). From the reported data about the classifications of the algorithms (i.e. true positives, false positives, false negatives, true negatives) a negative predictive value was calculated per algorithm. The negative predictive value of the algorithms varied between 0.69 and 1.00. See Table 1.5 for an overview per algorithm.

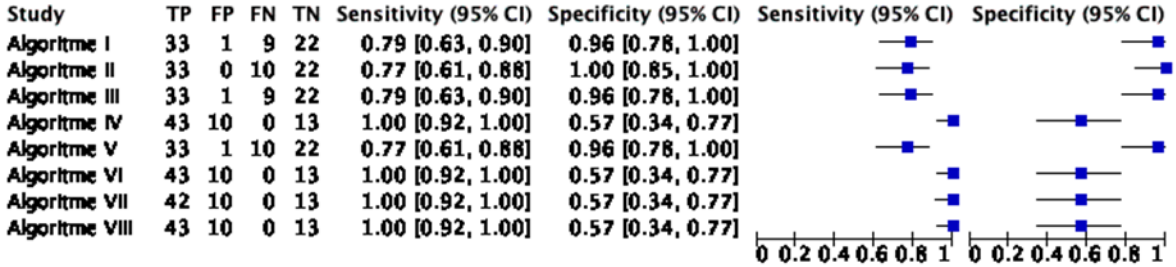
Table 1.5 An overview of the used modalities and the diagnostic accuracy per algorithm, in Van Cann (2008). The PPV, NPV, sensitivity, and specificity were calculated from the reported (mis)classifications

Algorithm	Sequential tests	Sample size	TP	FP	FN	TN	PPV	NPV	Sensitivity (95%CI)	Specificity (95%CI)
Algorithm I	SPECT-MRI-CT	65	33	1	9	22	0,97	0,71	0.79 (0.63-0.90)	0.96 (0.78-1.00)

Algorithm II	SPECT-CT-MRI	66	33	0	10	22	1,00	0,69	0.77 (0.61-0.88)	1.00 (0.85-1.00)
Algorithm III	MRI-SPECT-CT	65	33	1	9	22	0,97	0,71	0.79 (0.63-0.90)	0.96 (0.78-1.00)
Algorithm IV	MRI-CT-SPECT	66	43	10	0	13	0,81	1.00	1.00 (0.92-1.00)	0.57 (0.34-0.77)
Algorithm V	CT-SPECT-MRI	66	33	1	10	22	0,97	0,69	0.77 (0.61-0.88)	0.96 (0.78-1.00)
Algorithm VI	CT-MRI-SPECT	66	43	10	0	13	0,81	1,00	1.00 (0.92-1.00)	0.57 (0.34-0.77)
Algorithm VII	MRI-SPECT	65	42	10	0	13	0,80	1,00	1.00 (0.92-1.00)	0.57 (0.34-0.77)
Algorithm VIII	CT-SPECT	66	43	10	0	13	0,81	1,00	1.00 (0.92-1.00)	0.57 (0.34-0.77)

TP: True positive
FP: False positive
FN: False negative
TN: True negative
PPV: Positive predictive value
NPV: Negative predictive value
CI: Confidence interval

Figure 1.6 A graphical representation of the diagnostic accuracy per algorithm, in Van Cann (2008). Sensitivity and specificity were calculated from the reported (mis)classifications



<u>Level of evidence of the literature</u>

Modalities for mandibular bone invasion

<u>OPG</u>

Sensitivity (mandibular bone invasion)

The level of evidence regarding sensitivity (for mandibular bone invasion) was downgraded by 2 levels because of the study limitations (1 level for risk of bias: 72% of the judgements were 'unclear' in the assessment with the QUADAS-2 tool) and the number of included patients (1 level for imprecision: the upper and lower limit of the confidence interval of the pooled accuracy estimate may lead to different conclusions); Publication bias was not assessed.

Specificity (mandibular bone invasion):

The level of evidence regarding specificity (for mandibular bone invasion) was downgraded by 2 levels because of the study limitations (1 level for risk of bias: 72% of the judgements were 'unclear' in the assessment with the QUADAS-2 tool); Publication bias was not assessed.

Positive predictive value:

The positive predictive value was not reported and therefore a GRADE assessment could not be performed.

Negative predictive value:
The negative predictive value was not reported and therefore a GRADE assessment could not be performed.

CBCT
Sensitivity (mandibular bone invasion)
The level of evidence regarding sensitivity (for mandibular bone invasion) was downgraded by 2 levels because of the study limitations (1 level for risk of bias: 4 out of 5 judgements for the selection of participants, 3 out of 5 judgements for the reference test, and 3 out of 5 judgements for the flow and timing were 'unclear' in the assessment with the QUADAS-2 tool) and the number of included patients (1 level for imprecision: the number of included patients is relatively low); Publication bias was not assessed.

Specificity (mandibular bone invasion)
The level of evidence regarding specificity (for mandibular bone invasion) was downgraded by 2 levels because of the study limitations (1 level for risk of bias: 4 out of 5 judgements for the selection of participants, 3 out of 5 judgements for the reference test, and 3 out of 5 judgements for the flow and timing were 'unclear' in the assessment with the QUADAS-2 tool) and conflicting results (1 level for inconsistency: the heterogeneity could not be explained by the variables in the meta-regression for CBCT performed by Qiao (2018), The I^2 was 90%); Publication bias was not assessed.

Positive predictive value:
The positive predictive value was not reported and therefore a GRADE assessment could not be performed.

Negative predictive value:
The negative predictive value was not reported and therefore a GRADE assessment could not be performed.

CT
Sensitivity (mandibular bone invasion)
The level of evidence regarding sensitivity (for mandibular bone invasion) was downgraded by 2 levels because of the study limitations (1 level for risk of bias: 61% of the judgements were 'unclear' in the assessment with the QUADAS-2 tool) and conflicting results (1 level for inconsistency: the heterogeneity could not be explained by the variables in the meta-regression for CT performed by Qiao (2018), The I^2 was 70%); Publication bias was not assessed.

Specificity (mandibular bone invasion)
The level of evidence regarding specificity (for mandibular bone invasion) was downgraded by 1 level because of the study limitations (1 level for risk of bias: 61% of the judgements were 'unclear' in the assessment with the QUADAS-2 tool); Publication bias was not assessed.

Sensitivity (medullary bone invasion)
The level of evidence regarding sensitivity (for medullary bone invasion) was downgraded by 3 levels because of the study limitations (1 level for risk of bias: 65% of the judgements were 'unclear' in the assessment with the QUADAS-2 tool) and conflicting results (2 levels for inconsistency: the confidence intervals overlap insufficiently while the heterogeneity was unexplained, the I^2 was 83%); Publication bias was not assessed.

Specificity (medullary bone invasion)
The level of evidence regarding specificity (for medullary bone invasion) was downgraded by 3 levels because of the study limitations (1 level for risk of bias: 65% of the judgements were 'unclear' in the assessment with the QUADAS-2 tool) and the number of included patients (2 levels for imprecision: the number of patients was low); Publication bias was not assessed.

Positive predictive value
The positive predictive value was not reported and therefore a GRADE assessment could not be performed.

Negative predictive value
The negative predictive value was not reported and therefore a GRADE assessment could not be performed.

PET-CT
Sensitivity (mandibular bone invasion)
The level of evidence regarding sensitivity was downgraded by 3 levels because of the study limitations (1 level for risk of bias: Three of the four included studies had an unclear risk in the 'patient selection' and 'reference test' domains of the QUADAS-2 tool) and the number of included patients (2 levels for imprecision: the number of patients was low); Publication bias was not assessed.

Specificity (mandibular bone invasion)
The level of evidence regarding specificity was downgraded by 3 levels because of the study limitations (1 level for risk of bias: Three of the four included studies had an unclear risk in the 'patient selection' and 'reference test' domains of the QUADAS-2 tool) and the number of included patients (2 levels for imprecision: the number of patients was low); Publication bias was not assessed.

Positive predictive value
The positive predictive value was not reported and therefore a GRADE assessment could not be performed.

Negative predictive value
The negative predictive value was not reported and therefore a GRADE assessment could not be performed.

SPECT
Sensitivity (mandibular bone invasion)
The level of evidence regarding the sensitivity (for mandibular bone invasion) was downgraded by 1 level because of the study limitations (1 level for risk of bias: 57% of the risk of bias judgements were 'unclear' in the assessment with the QUADAS-2 tool); publication bias was not assessed.

Specificity (mandibular bone invasion)
The level of evidence regarding specificity (for mandibular bone invasion) was downgraded by 2 levels because of the study limitations (1 level for risk of bias: 57% of the risk of bias judgements were 'unclear' in the assessment with the QUADAS-2 tool) and conflicting results (1 level for inconsistency: the heterogeneity could not be explained by the variables in the meta-regression for SPECT performed by Qiao (2018)); Publication bias was not assessed.

Positive predictive value
The positive predictive value was not reported and therefore a GRADE assessment could not be performed.

Negative predictive value
The negative predictive value was not reported and therefore a GRADE assessment could not be performed.

Algorithms for mandibular bone invasion
Algorithms in Van Cann (2008)
Sensitivity
The level of evidence regarding sensitivity was downgraded by 3 levels because of the study limitations (1 level for risk of bias: judgements were largely 'unclear' in the risk of bias assessment) and the number of included patients (2 levels for imprecision: the number of included patients is very low); Publication bias was not assessed.

Specificity
The level of evidence regarding specificity was downgraded by 3 levels because of the study limitations (1 level for risk of bias: judgements were largely 'unclear' in the risk of bias assessment) and the number of included patients (2 levels for imprecision: the number of included patients is very low and the upper and lower limit of the confidence interval of the pooled accuracy estimate may lead to different conclusions); Publication bias was not assessed.

Positive predictive value
The level of evidence regarding positive predictive value was downgraded by 3 levels because of the study limitations (1 level for risk of bias: judgements were largely 'unclear' in the risk of bias assessment) and the number of included patients (2 levels for imprecision: the number of included patients is very low); Publication bias was not assessed.

Negative predictive value
The level of evidence regarding negative predictive value was downgraded by 3 levels because of the study limitations (1 level for risk of bias: judgements were largely 'unclear' in the risk of bias assessment) and the number of included patients (2 levels for imprecision: the number of included patients is very low); Publication bias was not assessed.

Conclusions
Imaging modalities (mandibular bone invasion)
The pooled accuracy estimates of imaging modalities and their certainty (following GRADE) for detecting mandibular bone invasion are presented in Table 1.7. Positive and negative predictive values were not reported.

Table 1.7 Pooled accuracy estimates and GRADE certainty for imaging modalities detecting mandibular bone invasion

Imaging modality	Sensitivity (95%CI) (GRADE certainty)	Specificity (95%CI) (GRADE certainty)	Positive / negative predictive values (95%CI) (GRADE certainty)	Reference
OPG	0.75 (0.67-0.82) (LOW)	0.83 (0.79-0.86) (MODERATE)	NA	*Qiao 2018*
CBCT	0.90 (0.85-0.93) (LOW)	0.85 (0.62-0.95) (LOW)	NA	*Qiao 2018*
CT	0.73 (0.66-0.80) (LOW)	0.91 (0.88-0.94) (MODERATE)	NA	*Qiao 2018*

PET-CT	0.90 (0.58-0.85) (VERY LOW)	0.89 (0.77-0.96) (VERY LOW)	NA	*Qiao 2018*
SPECT	0.97 (0.92-0.99) (MODERATE)	0.69 (0.52-0.82) (LOW)	NA	*Qiao 2018*
DECT	NA	NA	NA	*NA*
MRI	0.88 (0.78-0.94) (LOW)	0.90 (0.80-0.95) (LOW)	NA	*Qiao 2018*

Imaging modalities (medullary bone invasion)
The pooled accuracy estimates of imaging modalities and their certainty (following GRADE) for detecting medullary bone invasion are presented in Table 1.8. Positive and negative predictive values were not reported. The working-group decided not to present the test performance in a hypothetical cohort based on an arbitrarily chosen pretest probability.

Table 1.8 Pooled accuracy estimates and GRADE certainty for imaging modalities detecting medullary bone invasion

Imaging modality	Sensitivity (95%CI) (GRADE certainty)	Specificity (95%CI) (GRADE certainty)	Positive/negative predictive values (95%CI) (GRADE certainty)	Reference
OPG	NA	NA	NA	*NA*
CBCT	NA	NA	NA	*NA*
CT	0.85 (0.43-0.98) (VERY LOW)	0.86 (0.73-0.93) (VERY LOW)	NA	*Qiao 2018*
PET-CT	NA	NA	NA	*NA*
SPECT	NA	NA	NA	*NA*
DECT	NA	NA	NA	*NA*
MRI	0.93 (0.81-0.98) (VERY LOW)	0.84 (0.60-0.95) (VERY LOW)	NA	*Qiao 2018*

<u>Algorithms to detect mandibular bone invasion</u>

Very low GRADE	There is a very low certainty about the **sensitivity** of the diagnostic algorithm in Van Cann (2008) for diagnosing mandibular bone invasion by a tumor in patients with an oral cavity carcinoma. *Sources: (Van Cann, 2008)*

Very low GRADE	There is a very low certainty about the **specificity** of the diagnostic algorithm in Van Cann (2008) for diagnosing mandibular bone invasion by a tumor in patients with an oral cavity carcinoma. *Sources: (Van Cann, 2008)*

Very low GRADE	There is a very low certainty about the **positive predictive value** of the diagnostic algorithm in Van Cann (2008) for diagnosing mandibular bone invasion by a tumor in patients with an oral cavity carcinoma. *Sources: (Van Cann, 2008)*

Very low GRADE	There is a very low certainty about the **negative predictive value** of the diagnostic algorithm in Van Cann (2008) for diagnosing mandibular bone invasion by a tumor in patients with an oral cavity carcinoma. *Sources: (Van Cann, 2008)*

Overwegingen - van bewijs naar aanbeveling
Er is een redelijk vertrouwen in de gevonden sensitiviteit van SPECT (0,97; 95%BHI 0,92 tot 0,99) bij de diagnostisering van mandibulaire botinvasie. De zekerheid over de gevonden sensitiviteit was laag voor CBCT (0,90; 95%BHI 0,85 to 0,93), CT (0,73; 95%BHI 0,66 tot 0,80), MRI (0,88; 95%BHI 0,78 tot 0,94) en OPG (0,75; 95%BHI 0,67 tot 0,82) bij de diagnostisering van botinvasie en zeeg laag voor CT (0,85; 95%BHI 0,43 tot 0,98), PET-CT (0,90; 95%BHI: 0,58 tot 0.98) en MRI (0,93; 95%BHI 0,81 tot 0,98) bij de diagnostisering van beenmerginvasie door een tumor. Er werden geen data gerapporteerd voor DECT.

Voor de specificiteit van beeldvormende modaliteiten is er een redelijk vertrouwen in CT (0,91; 95%BHI 0,88 tot 0,94) en OPG (0,83; 95%BHI 0,79 tot 0,86) bij de diagnostisering van botinvasie. De zekerheid van de gevonden specificiteit van SPECT (0,69; 95%BHI 0,52 tot 0,82), CBCT (0,85; 95%BHI 0,62 tot 0,95), en MRI (0,90; 95%BHI 0,80 tot 0,95) was laag bij de diagnostisering van botinvasie en zeer laag voor CT (0,86; 95%BHI 0,73 tot 0,93), PET-CT (0,89; 95%BHI: 0,85 tot 1,00) en MRI (0,84; 95%BHI 0,60 tot 0,95) bij de diagnostisering van beenmerginvasie. Er werden geen data gerapporteerd voor DECT.

Er was tevens een zeer laag vertrouwen in de acht algoritmen uit Van Cann (2008), voornamelijk door het lage aantal deelnemers in de studie. De diagnostische test accuratesse van elk algoritme is te zien in Tabel 1.5 (tabblad 'Samenvatting literatuur'). De auteurs concludeerden dat het uitvoeren van een CT of MRI gevolgd door een SPECT het aantal mandibulaire resecties aanzienlijk zou verminderen en dat SPECT alleen noodzakelijk is indien de voorafgaande CT of MRI geen mandibulaire invasie laat zien (Van Cann, 2008).

Het zeer lage vertrouwen in beeldvormende modaliteiten voor het diagnosticeren van beenmerginvasie door tumoren wordt met name veroorzaakt door onverklaarbare heterogeniteit tussen studies en/of het lage aantal deelnemers in de studies (Qiao, 2018). Er werden geen data gerapporteerd met betrekking tot de positief en negatief voorspellende waarde in de systematische review van Qiao (2018). Na de zoekdatum van deze richtlijnmodule verscheen er een artikel over de diagnostische accuratesse van DECT voor het detecteren van beenmergoedeem (Timmer, 2020). Beenmergoedeem kan optreden bij botinvasie, maar ook bij trauma, bloedingen of ontstekingen. DECT (2^{e} of 3^{e} generatie scanners) werd als indextest afgezet tegenover MRI (STIR of T2-weighted MRI met vetonderdrukking). De gemiddelde tijd tussen DECT en MRI was 9 dagen (SD: 11). Er werden 33 patiënten geselecteerd die tussen 2016 en 2018 zowel DECT als MRI ondergingen voor hoofd-hals abnormaliteiten. Indicaties in de steekproef waren hoofd-hals maligniteiten (n=27, waarvan n=15 voor mondholte maligniteiten), infecties (n=3), goedaardige tumor (n=1), en veranderingen na radiotherapie (n=2). Alle beelden werden door twee radiologen onafhankelijk van elkaar beoordeeld en waren geblindeerd voor de indicatie, klinische diagnose en andere patiëntgegevens. De auteurs rapporteerden de sensitiviteit (0,85), specificiteit (0,92), positief voorspellende waarde (0,94) en negatief voorspellende waarde (0,80), maar geen bijgaande 95% betrouwbaarheidsintervallen. De auteurs concludeerden dat, wanneer er contra-indicaties zijn voor het maken van een MRI, DECT de potentie heeft om als alternatief het beenmergoedeem te kunnen detecteren.

Bij een onjuiste diagnose van botinvasie kan onder- of overbehandeling plaatsvinden. Bij een fout-positieve bevinding wordt onterecht een (uitgebreidere) mandibularesectie verricht. Het hierna ontstane defect dient meestal gereconstrueerd te worden en bemoeilijkt dentale rehabilitatie. Bij een fout-negatieve bevinding zal bij de tumorresectie de resectierand positief zijn, waardoor adjuvante behandeling (heroperatie en/of (chemo)radiotherapie) met bijbehorende morbiditeit nodig is en de prognose waarschijnlijk negatief beïnvloed wordt. Er

zijn geen aanvullende argumenten vanuit andere groepen of interventies bekend. Het genoemde probleem bij fout-positieve bevinding en op basis hiervan marginale mandibularesectie is een groter probleem bij patiënten met een lage onderkaak. Bij deze patiënten ontstaat eerder een continuïteitsonderbreking, waardoor meestal een uitgebreidere reconstructie nodig is.

Preoperatief is het belangrijk om zo goed mogelijk geïnformeerd te zijn over eventuele botinvasie van een primair mondholtecarcinoom. Deze informatie wordt gebruikt bij het opstellen van het therapieplan, met name of een marginale dan wel segmentale mandibularesectie dient te worden verricht. De uitgebreidheid van de operatie hangt hier dus (gedeeltelijk) vanaf. De meeste van deze beeldvormend onderzoeken kunnen poliklinisch laagdrempelig gemaakt worden (bijvoorbeeld OPG en CBCT) of worden al gemaakt om de uitbreiding in de weke delen te beoordelen (bijvoorbeeld MRI). De sensitiviteit van de CBCT en MRI is hoog bij een acceptabele specificiteit, waardoor het risico op positieve of krappe resectieranden (met indicatie voor adjuvante behandeling en kans op slechtere overleving) laag is en het risico op een te uitgebreide behandeling (met risico op extra morbiditeit) acceptabel is.

Een CT of MRI wordt in de meeste centra standaard vervaardigd om de uitbreiding naar weke delen te bepalen bij mondholtecarcinomen. Deze scans hebben ook een waarde bij de diagnostiek naar botinvasie. Het gebruik van deze, reeds vervaardigde beeldvorming, zal dan geen belasting en geen extra kosten met zich meebrengen voor het detecteren van botinvasie. Een CBCT is vaak gemakkelijk met lage kosten beschikbaar en heeft ook een goede sensitiviteit. Hoewel een SPECT (botscintigrafie) de hoogste sensitiviteit lijkt te hebben, is dit onderzoek duurder en minder gemakkelijk toegankelijk dan de alternatief beschikbare beeldvormende technieken, die laagdrempelig en tegen lagere kosten kunnen worden vervaardigd. Daarnaast is de kwaliteit van beeldvorming voor (vrijwel) alle modaliteiten in de loop der jaren verbeterd. Dit is een ook in de toekomst een continu proces.

Alle onderzochte beeldvormende onderzoeken zijn voor iedereen bereikbaar. SPECT en DECT zijn minder gemakkelijk toegankelijk, terwijl OPG, CBCT, CT en MRI juist gemakkelijk voor iedere patiënt toegankelijk zijn. Een OPG en CBCT worden routinematig zeer frequent op de polikliniek Mondziekten, Kaak-, en Aangezichtschirurgie verricht. CT en MRI zijn standaard onderzoeken bij patiënten met hoofd-halskanker. Niet elke patiënt kan echter een MRI ondergaan door angst, claustrofobie of metalen voorwerpen in het lichaam. Daarnaast kunnen er restricties zijn voor nierpatiënten om een CT met contrastvloeistof te ondergaan. Zie hiervoor ook de algemene contra-indicaties voor MRI en CT. In deze situaties zou een CBCT alleen mogelijk ook voldoende kunnen zijn. Er zijn geen subgroepen van patiënten bekend met andere waarden en voorkeuren.

Aanbevelingen

Aanbeveling-1

Rationale van de aanbeveling: weging van argumenten voor en tegen de interventies

Een OPG wordt veelal standaard verricht voor focusonderzoek. De sensitiviteit (met name voor invasie van het beenmerg) lijkt te laag te zijn, zodat bij een negatieve uitslag aanvullend onderzoek verricht dient te worden. CT en/of MRI worden veelal standaard voor andere doeleinden gemaakt. De sensitiviteit en specificiteit van deze modaliteiten lijken acceptabel te zijn voor het detecteren van botinvasie, zonder extra kosten. SPECT lijkt het meest sensitief te zijn, maar is minder goed toegankelijk en is duurder. Een CBCT lijkt een hoge sensitiviteit en specificiteit voor het detecteren van botinvasie van de mandibula te hebben.

CBCT is een onderzoek dat laagdrempelig en tegen lage kosten vaak poliklinisch kan worden verricht omdat veel MKA poliklinieken CBCT apparatuur beschikbaar hebben.

Gebruik voor het detecteren van botinvasie de beeldvormende technieken (CT en/of MRI) die standaard gebruikt worden voor het bepalen van de uitbreiding van de primaire tumor.

Overweeg bij een onzekere CT of MRI uitslag een CBCT te maken bij alle mondholtecarcinomen die tegen de mandibula aangelegen zijn om botinvasie te diagnosticeren.

Overweeg het gebruik van SPECT-CT alleen in geselecteerde gevallen waar de uitslagen van CT/MRI/CBCT onzeker zijn en het uitsluiten van botinvasie van invloed is op het chirurgisch behandelplan.

Literatuur

Qiao X, Liu W, Cao Y, Miao C, Yang W, Su N, Ye L, Li L, Li C. Performance of different imaging techniques in the diagnosis of head and neck cancer mandibular invasion: A systematic review and meta-analysis. Oral Oncol. 2018 Nov;86:150-164. doi: 10.1016/j.oraloncology.2018.09.024. Epub 2018 Sep 25. PubMed PMID: 30409295.

Timmer VCML, Kroonenburgh AMJLV, Henneman WJP, Vaassen LAA, Roele ED, Kessler PAWH, Postma AA. Detection of Bone Marrow Edema in the Head and Neck With Dual-Energy CT: Ready for Clinical Use? AJR Am J Roentgenol. 2020 Apr;214(4):893-899. doi: 10.2214/AJR.19.21881. Epub 2020 Feb 11. PMID: 32045307.

Van Cann EM, Koole R, Oyen WJ, de Rooy JW, de Wilde PC, Slootweg PJ, Schipper M, Merkx MA, Stoelinga PJ. Assessment of mandibular invasion of squamous cell carcinoma by various modes of imaging: constructing a diagnostic algorithm. Int J Oral Maxillofac Surg. 2008 Jun;37(6):535-41. doi: 10.1016/j.ijom.2008.02.009. Epub 2008 Apr 10. PubMed PMID: 18406107.

Geldigheid en Onderhoud

Module[1]	Regiehouder(s)[2]	Jaar van autorisatie	Eerstvolgende beoordeling actualiteit richtlijn[3]	Frequentie van beoordeling op actualiteit[4]	Wie houdt er toezicht op actualiteit[5]	Relevante factoren voor wijzigingen in aanbeveling[6]

[1] Naam van de module
[2] Regiehouder van de module (deze kan verschillen per module en kan ook verdeeld zijn over meerdere regiehouders)
[3] Maximaal na vijf jaar
[4] (half)Jaarlijks, eens in twee jaar, eens in vijf jaar
[5] regievoerende vereniging, gedeelde regievoerende verenigingen, of (multidisciplinaire) werkgroep die in stand blijft
[6] Lopend onderzoek, wijzigingen in vergoeding/organisatie, beschikbaarheid nieuwe middelen

Bijlagen bij module 1.2

Kennislacunes
Wat is de diagnostische accuratesse van SPECT, DECT, CBCT, CT, PET-CT, MRI, een orthopantogram of diagnostische algoritmen met een postoperatief pathologisch assessment als referentiestandaard om preoperatief de (mate van de) mandibulaire botinvasie of beenmerginvasie van een tumor te bepalen bij patiënten met mondholtekanker?

Indicatoren
Geen.

Implementatieplan

Aanbeveling	Tijdspad voor implementatie: < 1 jaar, 1 tot 3 jaar of 3 tot 5 jaar	Verwacht effect op kosten	Randvoorwaarden voor implementatie (binnen aangegeven tijdspad)	Mogelijke barrières voor implementatie[1]	Te ondernemen acties voor implementatie[2]	Verantwoordelijken voor acties[3]	Overige opmerkingen
Gebruik voor het detecteren van botinvasie de beeldvormende technieken (OPG, CT en/of MRI) die standaard gebruikt worden voor het bepalen van de uitbreiding van de primaire tumor.	< 1 jaar	OPG wordt al gebruikt voor focusonderzoek CT en MRI worden al gebruikt voor tumor uitbreiding in weke delen	OPG, CT en MRI beschikbaar	Geen	Geen	Lokale werkgroep	
Overweeg bij een negatieve OPG uitslag en onzekere CT of MRI uitslag een CBCT te maken bij alle mondholtecarcinomen die tegen de mandibula aangelegen zijn om botinvasie te diagnosticeren	1 tot 3 jaar	CBCT geeft extra kosten welke beperkt zijn	CBCT beschikbaar	Beschikbaarheid CBCT	Beschikbaar maken CBCT	Lokale werkgroep	
Overweeg het gebruik van SPECT-CT	1 tot 3 jaar	SPECT-CT geeft extra					

(botscintigrafie) alleen in geselecteerde gevallen waar de uitslagen van CT/MRI/CBCT onzeker zijn en de noodzaak van botinvasie relevant is in de diagnostiek.		kosten welke substantieel zijn					

[1] Barrières kunnen zich bevinden op het niveau van de professional, op het niveau van de organisatie (het ziekenhuis) of op het niveau van het systeem (buiten het ziekenhuis). Denk bijvoorbeeld aan onenigheid in het land met betrekking tot de aanbeveling, onvoldoende motivatie of kennis bij de specialist, onvoldoende faciliteiten of personeel, nodige concentratie van zorg, kosten, slechte samenwerking tussen disciplines, nodige taakherschikking, et cetera.

[2] Denk aan acties die noodzakelijk zijn voor implementatie, maar ook acties die mogelijk zijn om de implementatie te bevorderen. Denk bijvoorbeeld aan controleren aanbeveling tijdens kwaliteitsvisitatie, publicatie van de richtlijn, ontwikkelen van implementatietools, informeren van ziekenhuisbestuurders, regelen van goede vergoeding voor een bepaald type behandeling, maken van samenwerkingsafspraken.

[3] Wie de verantwoordelijkheden draagt voor implementatie van de aanbevelingen, zal tevens afhankelijk zijn van het niveau waarop zich barrières bevinden. Barrières op het niveau van de professional zullen vaak opgelost moeten worden door de beroepsvereniging. Barrières op het niveau van de organisatie zullen vaak onder verantwoordelijkheid van de ziekenhuisbestuurders vallen. Bij het oplossen van barrières op het niveau van het systeem zijn ook andere partijen, zoals de NZA en zorgverzekeraars, van belang.

Evidencetabellen

Table of quality assessment for systematic reviews of diagnostic studies

Based on AMSTAR checklist (Shea, 2007; BMC Methodol 7: 10; doi:10.1186/1471-2288-7-10) and PRISMA checklist (Moher, 2009, PLoS Med 6: e1000097; doi:10.1371/journal. pmed1000097)

Research question:

Study **First author, year**	**Appropriate and clearly focused question?[1]** **Yes/no/unclear**	**Comprehensive and systematic literature search?[2]** **Yes/no/unclear**	**Description of included and excluded studies?[3]** **Yes/no/unclear**	**Description of relevant characteristics of included studies?[4]** **Yes/no/unclear**	**Assessment of scientific quality of included studies?[5]** **Yes/no/unclear**	**Enough similarities between studies to make combining them reasonable?[6]** **Yes/no/unclear**	**Potential risk of publication bias taken into account?[7]** **Yes/no/unclear**	**Potential conflicts of interest reported?[8]** **Yes/no/unclear**
Qiao 2018	No Reason: the question should have been more specific on the domain and outcomes. Although, the focus of the SR can be seen in the inclusioncriteria.	Yes Reason: transparent reporting, although the exact search strategy (search string) is not reported.	No Reason: Excluded studies were not referenced.	Yes Reason: provided in table 1	Yes Reason: performed with QUADAS-2	Unclear Reason: setting and place of diagnostic test is not reported in the systematic review. Diagnostic criteria (or thresholds) were not reported. The authors do perform a meta-regression to assess heterogeneity.	No Reason: Authors state that publication bias was not assessed.	No Reason: Authors declare that there is no conflict of interest, however this is not reported for the individual studies.

1. **Research question (PICO) and inclusion criteria should be appropriate (in relation to the research question to be answered in the clinical guideline) and predefined.**
2. **Search period and strategy should be described; at least Medline searched.**
3. **Potentially relevant studies that are excluded at final selection (after reading the full text) should be referenced with reasons.**
4. **Characteristics of individual studies relevant to the research question (PICO) should be reported.**
5. **Quality of individual studies should be assessed using a quality scoring tool or checklist (preferably QUADAS-2; COSMIN checklist for measuring instruments) and taken into account in the evidence synthesis.**
6. **Clinical and statistical heterogeneity should be assessed; clinical: enough similarities in patient characteristics, diagnostic tests (strategy) to allow pooling? For pooled data: at least 5 studies available for pooling; assessment of statistical heterogeneity and, more importantly (see Note), assessment of the reasons for heterogeneity (if present)? Note: sensitivity and specificity depend on the situation in which the test is being used and the thresholds that have been set, and sensitivity and specificity are correlated; therefore, the use of heterogeneity statistics (p-values; I^2) is problematic, and rather than testing whether heterogeneity is present, heterogeneity should be assessed by eye-balling (degree of overlap of confidence intervals in Forest plot), and the reasons for heterogeneity should be examined.**

7. **There is no clear evidence for publication bias in diagnostic studies, and an ongoing discussion on which statistical method should be used. Tests to identify publication bias are likely to give false-positive results, among available tests, Deeks' test is most valid. Irrespective of the use of statistical methods, you may score "Yes" if the authors discuss the potential risk of publication bias.**
8. **Sources of support (including commercial co-authorship) should be reported in both the systematic review and the included studies. Note: To get a "yes," source of funding or support must be indicated for the systematic review AND for each of the included studies.**

Evidence table for systematic reviews of diagnostic test accuracy studies

Study reference	Study characteristics	Patient characteristics	Index test (test of interest)	Reference test	Follow-up	Outcome measures and effect size	Comments
Qiao 2018 (individual study characteristics deduced from Qiao 2018)	SR (and meta-analysis) *Literature searches up to November 2017.* 1: Abd El-Hafez 2011 2: Acton 2000 3: Ahuja 1990 4: Babin 2008 5: Bahadur 1990 6: Bolzoni 2004 7: Brockenbourgh 2003 8: Brown 1994 9: Chung 1994 10: Curran 1996 11: Dreiseidler 2011 12: Duan 2008 13: Gilbert 1986 14: Goerres 2005 15: Gu 2010 16: Hakim 2014 17: Handschel 2012	Inclusion criteria SR: DTA studies designed as cohort studies, patients diagnosed with oral cancer or head/neck cancer with preoperative biopsy and mandibulectomy during surgery, CT/ MRI / CBCT / OPG / PET-CT / SPECT / BS / US imaging as index test, pathological diagnosis as reference test, mandible invasion by the tumor as the target condition, TP / FP / FN / TN / Se / Sp / LR+ / LR- as outcomes. Exclusion criteria SR: Not stated *49 studies included.* <u>Important patient characteristics</u>: <u>Number of patients, mandibular invasion/total (prevalence)</u> 1: 37/114 (32.5%) 2: 36/67 (53.7%) 3: 30/48 (62.5%) 4: 3/17 (17.6%) 5: 11/44 (25%) 6: 14/43 (32.6%)	Describe index and comparator tests* and cut-off point(s): 1: PET-CT, MRI 2: SPECT, OPG, CT 3: OPG, BS 4: PET-CT, CT 5: CT, BS 6: MRI 7: CT 8: OPG, BS, CT, MRI 9: MRI 10: SOECT, CT 11: SPECT, CT, CBCT 12: SPECT, CT 13: BS	Describe reference test and cut-off point(s): 1-49: pathological diagnosis Prevalence (%) 1: 32.5% 2: 53.7% 3: 62.5% 4: 17.6% 5: 25% 6: 32.6% 7: 61.1% 8: 60% 9: 54.4% 10: 31% 11: 62.3% 12: 95.2% 13: 33.7% 14: 35.3% 15: 26.1% 16: 40% 17: 41.1%	Endpoint of follow-up: N/A	Outcome measures and effect size (include 95%CI and p-value if available): <u>SPECT sensitivity for mandibular invasion (95%CI):</u> 2: 0.60 (0.15-0.95) 10: 1.00 (0.66-1.00) 11: 0.92 (0.80-0.98) 12: 1.00 (0.83-1.00) 14: 0.92 (0.62-1.00) 22: 0.95 (0.75-1.00) 25: 1.00 (0.88-1.00)	<u>Study quality (ROB):</u> QUADAS-2 Authors performed meta-regression analysis as a means for heterogeneity analysis.

	18: Hendrikx 2010 19: Heppt 1993 20: Huang 2011 21: Imaizumi 2006 22: Imola 2001 23: Kalavrezos 1996 24: Kim 2013 25: Kolk 2014 26: Kushraj 2011 27: Lane 2000 28: Leipzig 1985 29: Linz 2015 30: Luyk 1986 31: Millesi 1990 32: Momin 2009 33: Mukherji 2001 34: Ord 1997 35: Rajesh 2008 36: Rao 2004 37: Schimming 2000 38: Smyth 1996 39: Soderholm 1990 40: Suzuki 2004 41: Tsue 1994 42: Van Cann 2008a 43: Van Cann 2008b 44: Van den Brekel 1998 45: Vidiri 2010 46: Wiener 2006 47: Yamamoto 2002 48: Zieron 2001	7: 22/36 (61.1%) 8: 21/35 (60%) 9: 12/22 (54.4%) 10: 9/29 (31%) 11: 48/77 (62.3%) 12: 20/21 (95.2%) 13: 35/104 (33.7%) 14: 12/34 (35.3%) 15: 12/46 (26.1%) 16: 84/210 (40%) 17: 44/107 (41.1%) 18: 11/23 (47.8%) 19: 13/33 (39.4%) 20: 8/17 (47.1%) 21: 25/51 (49%) 22: 20/38 (52.6%) 23: 48/60 (80%) 24: 20/27 (74.1%) 25: 28/50 (56%) 26: 8/15 (53.3%) 27: 14/26 (53.8%) 28: 17/31 (54.8%) 29: 66/197 (33.5%) 30: 4/11 (36.4%) 31: 31/97 (32%) 32: 43/50 (86%) 33: 26/49 (53.1%) 34: 15/46 (32.6%) 35: 19/23 (82.6%) 36: 24/51 (47.1%) 37: 40/88 (45.5%) 38: 27/40 (67.5%) 39: 10/13 (76.9%) 40: 17/34 (50%) 41: 25/64 (39.1%) 42: 44/67 (65.7%) 43: 12/25 (48%)	14: SPECT, PET-CT, CT 15: CT, MRI, PET-CT, 16: CT, CBCT, BS, SPECT 17: CT 18: OPG, MRI, CBCT 19: US 20: PET-CT, MRI, CT 21: CT, MRI 22: SPECT, OPG, CT 23: BS, CT 24: MRI, PET-CT 25: SPECT, CT, MRI, OPG 26: OPG, CT, SPECT 27: CT 28: BS 29: OPG, CBCT, BS, CT, MRI 30: BS, OPG 31: CT, BS, US 32: OPG, CBCT 33: CT	18: 47.8% 19: 39.4% 20: 47.1% 21: 49% 22: 2.6% 23: 80% 24: 74.1% 25: 56% 26: 53.3% 27: 53.8% 28: 54.8% 29: 33.5% 30: 36.4% 31: 32% 32: 86% 33: 53.1% 34: 32.6% 35: 82.6% 36: 47.1% 37: 45.5% 38: 67.5% 39: 76.9% 40: 50% 41: 39.1% 42: 65.7% 43: 48% 44: 62.1% 45: 38.9% 46: 15.4% 47: 33.3% 48: 46.1% 49: 46% For how many participants were no complete		26: 1.00 (0.63-1.00) 35: 1.00 (0.82-1.00) 40: 0.95 (0.74-1.00) 42: 1.00 (0.92-1.00) 47: 1.00 (0.75-1.00) 48: 0.95 (0.83-0.99) Sensitivity (statistical model not reported): 0.97 (95%CI 0.92 to 0.99) Heterogeneity (reasons): I^2 = 72% <u>SPECT specificity for mandibular invasion (95%CI):</u> 2: 0.67 (0.30-0.93) 10: 0.29 (0.04-0.71) 11: 0.41 (0.24-0.61) 12: 0.00 (0.00-0.98) 14: 0.86 (0.65-0.97)	

	49: Zupi 1996 Study design: cohort (by inclusion criteria), retrospective / prospective 1: Retrospect. 2: Prospect. 3: Retrospect. 4: Prospect. 5: Retrospect. 6: Prospect. 7: Retrospect. 8: Prospect. 9: Retrospect. 10: Retrospect. 11: Prospect. 12: Retrospect. 13: Retrospect. 14: Prospect. 15: Retrospect. 16: Retrospect. 17: Retrospect. 18: Retrospect. 19: Prospect. 20: Prospect. 21: Retrospect. 22: Prospect. 23: Retrospect. 24: Prospect. 25: Prospect. 26: Prospect. 27: Retrospect. 28: Prospect. 29: Retrospect.	44: 18/29 (62.1%) 45: 14/36 (38.9%) 46: 8/52 (15.4%) 47: 13/39 (33.3%) 48: 41/89 (46.1%) 49: 23/50 (46%) *Mean age in years, Sex as % male:* 1: NR (NR) 2: 61.6 (70.1%) 3: NR (NR) 4: NR (88.2%) 5: range: 22-65 (70.5%) 6: 57 (86%) 7: 65.4 (63.9%) 8: 64.9 (80%) 9: NR (NR) 10: 67 (72.4%) 11: 61 (67.5%) 12: 53.2 (85.7%) 13: 61 (75%) 14: 64.2 (50%) 15: 59.4 (84.8%) 16: NR (NR) 17: 62 (NR) 18: 63 (NR) 19: NR (NR) 20: 54 (94.1%) 21: 61 (76.5%) 22: 60.7 (76.3%) 23: 60.2 (73.3%) 24: 73.6 (40.7%) 25: 61 (86%) 26: NR (NR) 27: NR (NR) 28: NR (NR)	34: CT, OPG 35: PET-CT, CT, MRI 36: OPG 37: OPG, CT, SPECT 38: OPG, CT, MRI 39: BS, OPG 40: SPECT, CT 41: CT, MRI 42: SPECT, CT, MRI 43: CT, MRI 44: OPG, CT, MRI 45: CT, MRI 46: CT, MRI 47: SPECT, CT 48: SPECT, CT 49: BS, MRI, CT	outcome data available? N (%) Not reported Reasons for incomplete outcome data described. Not reported		22: 0.72 (0.47-0.90) 25: 1.00 (0.85-1.00) 26: 0.14 (0.00-0.58) 35: 0.50 (0.07-0.93) 37: 0.92 (0.80-0.98) 40: 0.73 (0.45-0.92) 42: 0.57 (0.34-0.77) 47: 0.85 (0.65-0.96) 48: 0.56 (0.41-0.71) Specificity (statistical model not reported): 0.69 (95% CI 0.52 to 0.82) Heterogeneity (reasons): I^2 = 79% SPECT AUC for mandibular invasion (95%CI): AUC (SE) from HSROC curve: AUC = 0.943 (0.039)	

	30: Retrospect. 31: Prospect. 32: Prospect. 33: Retrospect. 34: Retrospect. 35: Retrospect. 36: Prospect. 37: Prospect. 38: Retrospect. 39: Prospect. 40: Prospect. 41: Retrospect. 42: Prospect. 43: Prospect. 44: Retrospect. 45: Retrospect. 46: Retrospect. 47: Prospect. 48: Retrospect. 49: Retrospect. <u>Country</u>: 1: China 2: Asutralia 3: Scotland 4: France 5: India 6: Italia 7: US 8: UK 9: US 10: Ireland 11: Germany 12: China 13: US 14: Sweden	29: 63.7 (66.5%) 30: range:42-74 (63.6%) 31: NR (NR) 32: 55 (60%) 33: 59 (69.4%) 34: 63.2 (56.5%) 35: NR (NR) 36: 53.4 (58.8%) 37: 51.5 (83%) 38: 57 (82.5%) 39: range:40-89 (53.9%) 40: 63 (67.6%) 41: 62 (50%) 42: 63 (62.7%) 43: 54 (60%) 44: 57 (65.5%) 45: 56 (72.2%) 46: 63 (67.3%) 47: 63.2 (53.8%) 48: NR (NR) 49: NR (56%) <u>Types of tumors included:</u> A: cheek B: gingiva C: tongue D: floor of mouth E: retromolar trigone F: palate G: oral cavity H: oropharynx I: lymph node J: lip K: tonsil L: submandibular triangle M: submandibular gland N: mandible				<u>CBCT sensitivity for mandibular invasion (95%CI):</u> 11: 0.92 (0.80-0.98) 16: 0.94 (0.79-0.99) 18: 0.91 (0.59-1.00) 29: 0.88 (0.78-0.95) 32: 0.88 (0.75-0.96) Sensitivity (statistical model not reported): 0.90 (95% CI 0.85 to 0.93) Heterogeneity (reasons): I^2 = 0% <u>CBCT specificity for mandibular invasion (95%CI):</u> (e.g., sensitivity / specificity (%))	

	15: Korea 16: Germany 17: Germany 18: Netherlands 19: Germany 20: China 21: Japan 22: US 23: Sweden 24: Japan 25: Germany 26: India 27: US 28: US 29: Germany 30: UK 31: Australia 32: Japan 33: US 34: US 35: UK 36: India 37: Germany 38: Ireland 39: Finland 40: Japan 41: USA 42: Netherlands 43: Netherlands 44: Netherlands 45: Italia 46: Germany 47: Japan 48: Germany 49: Italia	O: pharynx 1: A, B, C, F, E, F 2: D, B, E 3: G 4: G, H 5: A, D, C, B, E, K 6: G, H 7: G 8: D, B, E, I 9: G, H 10: G, H 11: G 12: B 13: C, K, E, B, D 14: B, E 15: G 16: G 17: D, C, B 18: E, D, B 19: D, K 20: A 21: B, D, A 22: B, E, C, K 23: G 24: G 25: B, D, C, F, A 26: G 27: E 28: G 29: D, J, F, N, C, A 30: B, D, L 31: D, C 32: B 33: G 34: D, E, A, B 35: G				11: 0.97 (0.82-1.00) 16: 0.59 (0.39-0.78) 18: 1.00 (0.74-1.00) 29: 0.83 (0.76-0.89) 32: 0.57 (0.18-0.90) Specificity (statistical model not reported): 0.85 (95% CI 0.62 to 0.95) Heterogeneity (reasons): I^2 = 80% <u>CBCT AUC for mandibular invasion (95%CI):</u> AUC (SE) from HSROC curve: AUC = 0.946 (0.015) <u>CT sensitivity for mandibular invasion (95%CI):</u> 2: 0.67 (0.46-0.83)	

	Source of funding and conflicts of interest: Authors state that there is no CoI. Not reported for individual studies.	36: G 37: C, E, B, D 38: D, B, K 39: G 40: N, D, C, E, J, A, M 41: B, E, D, J, C, A, K, H 42: D, E, B, A 43: D, B, E 44: D, E 45: G 46: B, D, C, F, A 47: C, A, B, D 48: B, D, C, O 49: G				4: 0.33 (0.01-0.91) 5: 0.80 (0.28-0.99) 7: 0.95 (0.77-1.00) 8: 0.42 (0.15-0.72) 10: 0.89 (0.52-1.00) 11: 0.79 (0.65-0.90) 12: 0.80 (0.56-0.94) 14: 0.92 (0.62-1.00) 15: 0.42 (0.15-0.72) 16: 0.63 (0.45-0.79) 17: 0.82 (0.67-0.92) 20: 0.56 (0.21-0.86) 21: 1.00 (0.86-1.00) 22: 0.55 (0.32-0.77) 23: 0.78 (0.60-0.91) 25: 0.89 (0.72-0.98) 26: 0.75 (0.35-0.97) 27: 0.50 (0.23-0.77)	

						29: 0.64 (0.48-0.78) 33: 0.96 (0.80-1.00) 34: 0.53 (0.27-0.79) 35: 0.95 (0.74-1.00) 37: 0.90 (0.76-0.97) 38: 0.63 (0.35-0.85) 40: 0.71 (0.44-0.90) 41: 0.50 (0.28-0.72) 42: 0.58 (0.42-0.73) 43: 0.38 (0.18-0.62) 44: 0.64 (0.35-0.87) 45: 0.79 (0.49-0.95) 46: 0.75 (0.34-0.97) 47: 0.38 (-.14-0.68) 48: 0.63 (0.41-0.81) 49: 0.91 (0.72-0.99) Sensitivity (statistical model not reported):	

						0.73 (95% CI 0.66 to 0.80) Heterogeneity (reasons): I2 = 70% <u>CT specificity for mandibular invasion (95%CI):</u> 2: 0.90 (0.76-0.97) 4: 1.00 (0.77-1.00) 5: 0.94 (0.71-1.00) 7: 0.79 (0.49-0.95) 8: 1.00 (0.69-1.00) 10: 0.75 (0.43-0.95) 11: 1.00 (0.88-1.00) 12: 1.00 (0.03-1.00) 14: 1.00 (0.85-1.00) 15: 1.00 (0.90-1.00) 16: 0.84 (0.69-0.93) 17: 0.87 (0.77-0.94) 20: 0.50 (0.16-0.84)	

						21: 0.88 (0.70-0.98) 22: 0.89 (0.65-0.99) 23: 0.80 (0.52-0.96) 25: 1.00 (0.85-1.00) 26: 1.00 (0.59-1.00) 27: 0.92 (0.62-1.00) 29: 0.86 (0.75-0.93) 33: 0.87 (0.66-0.97) 34: 0.93 (0.76-0.99) 35: 1.00 (0.40-1.00) 37: 0.94 (0.83-0.99) 38: 0.75 (0.53-0.90) 40: 1.00 (0.75-1.00) 41: 0.86 (0.67-0.96) 42: 0.96 (0.78-1.00) 43: 0.75 (0.19-0.99) 44: 0.89 (0.52-1.00) 45: 0.82 (0.60-0.95)	

						46: 0.95 (0.85-0.99) 47: 0.96 (0.80-1.00) 48: 0.78 (0.56-0.93) 49: 0.96 (0.81-1.00) Specificity (statistical model not reported): 0.91 (95% CI 0.88 to 0.94) Heterogeneity (reasons): I^2 = 49% <u>CT AUC for mandibular invasion (95%CI):</u> AUC (SE) from HSROC curve: AUC = 0.899 (0.029) <u>CT sensitivity for bone marrow invasion (95%CI):</u> 21: 1.00 (0.86-1.00) 38: 0.90 (0.55-1.00)	

						43: 0.67 (0.35-0.90) 44: 0.40 (0.12-0.74) Sensitivity (statistical model not reported): 0.85 (95% CI 0.43 to 0.98) Heterogeneity (reasons): I^2 = 83% <u>CT specificity for bone marrow invasion (95%CI):</u> 21: 0.88 (0.70-0.98) 38: 0.77 (0.58-0.90) 43: 1.00 (0.75-1.00) 44: 0.85 (0.55-0.98) Specificity (statistical model not reported): 0.86 (95% CI 0.73 to 0.93) Heterogeneity (reasons): I^2 = 32%	

						CT AUC for bone marrow invasion (95%CI): AUC (SE) from HSROC curve: AUC = 0.924 (0.045) PET-CT sensitivity for mandibular invasion (95%CI): 4: 1.00 (0.29-1.00) 14: 1.00 (0.74-1.00) 15: 0.58 (0.28-0.85) 20: 0.88 (0.47-1.00) Sensitivity (statistical model not reported): 0.90 (95% CI 0.58 to 0.85) Heterogeneity (reasons): I^2 = 64% PET-CT specificity for mandibular invasion (95%CI):	

						4: 0.86 (0.57-0.98) 14: 0.91 (0.71-0.99) 15: 0.97 (0.85-1.00) 20: 0.67 (0.30-0.93) <u>PET-CT AUC for mandibular invasion (95%CI):</u> AUC (SE) from HSROC curve: AUC = 0.929 (0.0349) <u>MRI sensitivity for mandibular invasion (95%CI):</u> 6: 0.93 (0.68-1.00) 8: 0.91 (0.59-1.00) 9: 1.00 (0.74-1.00) 15: 0.58 (0.28-0.85) 18: 0.82 (0.48-0.98) 20: 0.75 (0.35-0.97)	

						21: 0.96 (0.80-1.00) 25: 0.96 (0.82-1.00) 29: 0.75 (0.53-0.90) 35: 1.00 (0.82-1.00) 38: 0.67-0.09-0.99) 41: 1.00 (0.29-1.00) 42: 0.63 (0.47-0.77) 43: 0.58 (0.28-0.85) 44: 0.94 (0.73-1.00) 45: 0.93 (0.66-1.00) 46: 1.00 (0.63-1.00) 49: 0.39 (0.20-0.61) Sensitivity (statistical model not reported): 0.88 (95% CI 0.78 to 0.94) Heterogeneity (reasons): I^2 = 76% <u>MRI specificity for mandibular</u>	

						invasion (95%CI): 6: 0.93 (0.76-0.99) 8: 1.00 (0.29-1.00) 9: 0.40 (0.12-0.74) 15: 0.97 (0.85-1.00) 18: 0.67 (0.35-0.90) 20: 0.75 (0.35-0.97) 21: 0.54 (0.33-0.73) 25: 0.95 (0.77-1.00) 29: 0.97 (0.91-1.00) 35: 0.75 (0.19-0.99) 38: 1.00 (0.48-1.00) 41: 0.50 (0.16-0.84) 42: 1.00 (0.85-1.00) 43: 1.00 (0.75-1.00) 44: 0.73 (0.39-0.94) 45: 0.82 (0.60-0.95) 46: 0.93 (0.81-0.99)	

						49: 0.96 (0.81-1.00) Specificity (statistical model not reported): 0.90 (95% CI 0.80 to 0.95) Heterogeneity (reasons): I^2 = 81% <u>MRI AUC for mandibular invasion (95%CI):</u> AUC (SE) from HSROC curve: AUC = 0.929 (0.014) <u>MRI sensitivity for bone marrow invasion (95%CI):</u> 1: 0.97 (0.86-1.00) 6: 1.00 (0.69-1.00) 9: 1.00 (0.48-1.00) 21: 0.96 (0.80-1.00) 24: 0.95 (0.75-1.00)	

						43: 0.58 (0.28-0.85) 44: 0.83 (0.52-0.98) Sensitivity (statistical model not reported): 0.93 (95% CI 0.81 to 0.98) Heterogeneity (reasons): I^2 = 79% <u>MRI specificity for bone marrow invasion (95%CI):</u> 1: 0.61 (0.49-0.72) 6: 1.00 (0.89-1.00) 9: 0.71 (0.44-0.90) 21: 0.81 (0.61-0.93) 24: 0.57 (0.18-0.90) 43: 1.00 (0.75-1.00) 44: 0.65 (0.38-0.86) Specificity (statistical model not reported):	

						0.84 (95% CI 0.60 to 0.95) Heterogeneity (reasons): I^2 = 79% <u>MRI AUC for bone marrow invasion (95%CI):</u> AUC (SE) from HSROC curve: AUC = 0.934 (0.030) <u>OPG sensitivity for mandibular invasion (95%CI):</u> 2: 0.75 (0.58-0.88) 3: 0.93 (0.76-0.99) 8: 0.76 (0.53-0.92) 18: 0.55 (0.23-0.83) 22: 0.50 (0.27-0.73) 25: 0.79 (0.59-0.92) 26: 0.75 (0.35-0.97) 29: 0.59 (0.46-0.71)	

						30: 1.00 (0.40-1.00) 32: 0.72 (0.56-0.85) 34: 0.87 (0.60-0.98) 36: 0.92 (0.74-0.99) 37: 0.82 (0.67-0.93) 38: 0.81 (0.54-0.96) 39: 0.40 (0.12-0.74) Sensitivity (statistical model not reported): 0.75 (95% CI 0.67 to 0.82) Heterogeneity (reasons): I^2 = 62% <u>OPG specificity for mandibular invasion (95%CI):</u> 2: 0.71 (0.52-0.86) 3: 0.71 (0.48-0.89) 8: 0.92 (0.62-1.00) 18: 0.92 (0.62-1.00)	

						22: 0.94 (0.73-1.00) 25: 0.82 (0.60-0.95) 26: 1.00 (0.59-1.00) 29: 0.82 (0.60-0.95) 30: 1.00 (0.59-1.00) 32: 0.86 (0.42-1.00) 34: 0.80 (0.61-0.92) 36: 0.88 (0.70-0.98) 37: 0.79 (0.65-0.90) 38: 0.96 (0.79-1.00) 39: 0.67 (0.09-0.99) Specificity (statistical model not reported): 0.83 (95% CI 0.79 to 0.86) Heterogeneity (reasons): I^2 = 19% <u>OPG AUC for mandibular invasion (95%CI):</u>	

						AUC (SE) from HSROC curve: AUC = 0.876 (0.022) <u>OPG for bone marrow invasion (95%CI):</u> Only 1 study measured bone marrow invasion with OPG. From the systematic review it was unclear which study. Sensitivity: 0.63 (95%CI NR) Specificity: 0.90 (95%CI: NR) AUC = N/A	

***comparator test equals the C of the PICO; two or more index/ comparator tests may be compared; note that a comparator test is not the same as a reference test (golden standard)**

Voor de risk of bias en applicability assessment van individuele studies met het QUADAS-2 beoordelingsinstrument wordt verwezen naar de systematische review van Qiao (2018).

Risk of bias assessment diagnostic accuracy studies (QUADAS II, 2011) voor algoritmen

Study reference	Patient selection	Index test	Reference standard	Flow and timing	Comments with respect to applicability
Van Cann 2008	Was a consecutive or random sample of patients enrolled? Consecutive Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes	Were the index test results interpreted without knowledge of the results of the reference standard? Unclear (observers were unaware of results from other imaging modalities, however such statement was not found with regards to the reference test) If a threshold was used, was it pre-specified? Thresholds were implicit. Yes (for MRI and CT) No (for spect)	Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Unclear (not reported who assessed the specimens and whether they were blinded for index test results)	Was there an appropriate interval between index test(s) and reference standard? Unclear (time between index and reference test was not described) Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? No (there were some exclusions due to CT artefacts, uncertain MR imaging, uncertain SPECT imaging, it is unclear how this may have affected the results since it concerned a low number of exclusions in each analysis)	Are there concerns that the included patients do not match the review question? No Are there concerns that the index test, its conduct, or interpretation differ from the review question? No Are there concerns that the target condition as defined by the reference standard does not match the review question? No
	CONCLUSION: Could the selection of patients have introduced bias? **RISK: LOW**	CONCLUSION: Could the conduct or interpretation of the index test have introduced bias? **RISK: UNCLEAR**	CONCLUSION: Could the reference standard, its conduct, or its interpretation have introduced bias? **RISK: UNCLEAR**	CONCLUSION Could the patient flow have introduced bias? **RISK: Unclear**	

Judgments on risk of bias are dependent on the research question: some items are more likely to introduce bias than others, and may be given more weight in the final conclusion on the overall risk of bias per domain:

Patient selection:

- **Consecutive or random sample has a low risk to introduce bias.**
- **A case control design is very likely to overestimate accuracy and thus introduce bias.**
- **Inappropriate exclusion is likely to introduce bias.**

Index test:

- **This item is similar to "blinding" in intervention studies. The potential for bias is related to the subjectivity of index test interpretation and the order of testing.**
- **Selecting the test threshold to optimise sensitivity and/or specificity may lead to overoptimistic estimates of test performance and introduce bias.**

Reference standard:

- **When the reference standard is not 100% sensitive and 100% specific, disagreements between the index test and reference standard may be incorrect, which increases the risk of bias.**
- **This item is similar to "blinding" in intervention studies. The potential for bias is related to the subjectivity of index test interpretation and the order of testing.**

Flow and timing:

- **If there is a delay or if treatment is started between index test and reference standard, misclassification may occur due to recovery or deterioration of the condition, which increases the risk of bias.**
- **If the results of the index test influence the decision on whether to perform the reference standard or which reference standard is used, estimated diagnostic accuracy may be biased.**
- **All patients who were recruited into the study should be included in the analysis, if not, the risk of bias is increased.**

Judgement on applicability:

Patient selection: there may be concerns regarding applicability if patients included in the study differ from those targeted by the review question, in terms of severity of the target condition, demographic features, presence of differential diagnosis or co-morbidity, setting of the study and previous testing protocols.

Index test: if index tests methods differ from those specified in the review question there may be concerns regarding applicability.

Reference standard: the reference standard may be free of bias but the target condition that it defines may differ from the target condition specified in the review question.

Evidence table for diagnostic test accuracy studies

Study reference	Study characteristics	Patient characteristics	Index test (test of interest)	Reference test	Follow-up	Outcome measures and effect size	Comments
Van Cann 2008	Type of study: cohort (prospective data collection, retrospective analyses) Setting and country: University hospital, the Netherlands Funding and conflicts of interest: no CoI/funding statements reported in the manuscript	Inclusion criteria: histologically confirmed tumors fixed or adjacent to the mandible Exclusion criteria: history of surgery or radiotherapy in the head/neck area. Osteomyelitis, osteonecrosis, recent dental extraction, recent biopsy, trauma to the mandible N=67 Prevalence: 44/67 = 65.7% Mean age (range): 63 (43-84). Sex: 42/67M (62.7%) / 25F (37.2%) Other important characteristics: Tumor location: Floor of mouth: 31 Retromolar area: 20	Describe individual index tests: Clinical examinations (procedures not described) OPT (Periapical radiographs if possible due to pain or limited access) (radiographs were taken tangential to the suspected area. Two surgeons were informed of the SCC site and assessed the radiographs, but unaware of findings of other imaging modalities) SPECT (SPECT with a dual head gamma camera was performed 3-4 hours after injection with Tc-99m-methylene-diphosphonate. Two nuclear medicine physicians assessed the scans, unaware of findings from other imaging modes)	Describe reference test: Pathohistological assessment. Specimens were devided in two groups: Group C(M): present cortical bone invasion or present medullary invasion Group O: no bone invasion Cut-off point(s): Cortical invasion: replacement of bone by an advancing tumor front without invasion to cancellous spaces / dental canal / periodontal ligament Medullary invasion: diffuse growth through the cortex into cancellous bone / dental canal / periodontal ligament. No invasion: a continuous periosteal layer separating the tumor form bone.	Time between the index test and reference test: not described. For how many participants were no complete outcome data available? 10 participants were excluded because not all imaging data was complete: extreme obesity (n=1), claustrophobia (n=3), no slots available for imaging (n=6) 1 MRI scan and 1 SPECT scan were uncertain. Two CT scans showed artefacts due to metallic dental restorations. Reasons for incomplete outcome data described.	Outcome measures and effect size (include 95%CI and p-value if available): Algorithm I (2 CT artefacts, n=67) Start with SPECT. • Negative SPECT: 13/13 TN • Positive SPECT continue with MRI. ○ Negative MRI, continue with CT. ▪ Negative CT: 9/18 TN, 9/18 FN ▪ Positive CT: 6/7 TP, 1/7 FP ○ Positive MRI: 27/27 TP Total false results: 10 False positive: 0.0303 (95%CI: 0.000767-0.158) False negative: 0.290 (95%CI: 0.142-0.480) **Algorithm II** (1 uncertain MRI, n=67) Start with SPECT. • Negative SPECT: 13/13 TN • Positive SPECT, continue with CT. ○ Negative CT continue with MRI.	10 out of 77 participants were excluded because not all imaging data was complete: extreme obesity (n=1), claustrophobia (n=3), no slots available for imaging (n=6)

		Lower aveolar process: 13 Cheek mucosa: 3	CT (spiral CT was performed after injection of iohexol. 1.5mm slices were recorded in the area of suspected invasion. Two radiologists assessed the scans, unaware of findings from other modalities) MRI (Fast spin echo T2-weightet images were recorded on a 1.5T MR system. Two radiologists assessed the MR images, unaware of the findings from other modalities) Cut-off point(s): For CT: absence of cortex adjacent to an abnormal soft-tissue mass For MRI: replacement of hypointense signal of cortical bone by the signal intensity of tumour on both the SE T2-weightet and SE T1-weighted images -or- as a replacement of hyperintense signal of		Extreme obesity (n=1), claustrophobia (n=3), no slots available for imaging (n=6) Two CT scans had artefacts due to metallic dental restorations.	▪ Negative MRI: 9/19 TN, 10/19 FN ▪ Positive MRI: 8/8TP ○ Positive CT: 25/26 TP, 1/26 FP Total false results: 11 False positive: 0.0303 (95%CI: 0.000767-0.158) False negative: 0.313 (95%CI: 0.161-0.500) **Algorithm III** (2 CT artefacts, n=67) Start with MRI. • Positive MRI: 27/27TP • Negative MRI continue with SPECT. ○ Positive SPECT, continue with CT. ▪ Negative CT: 9/18 TN, 9/18 FN ▪ Positive CT: 6/7 TP, 1/7 FP ○ Negative SPECT: 13/13TN Total false results: 10 False positive: 0.0294 (95%CI: 0.000744-0.153) False negative: 0.290 (95%CI: 0.142-0.480) **Algorithm IV** (1 uncertain SPECT, n=67)	

			medullary bone by tumor signal intensity For other modalities: N/A Describe test strategies: Algorithm I (2 CT artefacts) Start with SPECT. • Negative SPECT • Positive SPECT continues with MRI. ○ Negative MRI, continue with CT. ▪ Negative CT ▪ Positive CT ○ Positive MRI Algorithm II (1 uncertain MRI) Start with SPECT. • Negative SPECT • Positive SPECT, continue with CT. ○ Negative CT continues with MRI. ▪ Negative MRI ▪ Positive MRI ○ Positive CT Algorithm III (2 CT artefacts) Start with MRI.			Start with MRI. • Positive MRI: 27/27TP • Negative MRI, continue with CT. ○ Negative CT continue with SPECT. ▪ Negative SPECT: 13/13 TN ▪ Positive SPECT: 10/19 TN, 9/10 FP ○ Positive CT: 6/7 TP, 1/7 TN Total false results: 10 False positive: 0.385 (95%CI: 0.202-0.594) False negative: 0 (95%CI: 0-0.0881) **Algorithm V** (1 uncertain MRI, n=67) Start with CT. • Positive CT: 25/26 TP, 1/26 FP • Negative CT continue with SPECT. ○ Positive SPECT continue with MRI. ▪ Negative MRI: 9/19 TN, 10/19 FN ▪ Positive MRI: 8/8 TP ○ Negative SPECT: 13/13 TN Total false results: 11	

			• Positive MRI • Negative MRI continues with SPECT. ○ Positive SPECT, continue with CT. ▪ Negative CT ▪ Positive CT ○ Negative SPECT Algorithm IV (1 uncertain SPECT) Start with MRI. • Positive MRI • Negative MRI, continue with CT. ○ Negative CT continues with SPECT. ▪ Negative SPECT ▪ Positive SPECT ○ Positive CT Algorithm V (1 uncertain MRI) Start with CT. • Positive CT • Negative CT continues with SPECT. ○ Positive SPECT continues with MRI. ▪ Negative MRI ▪ Positive MRI ○ Negative SPECT			False positive: 0.0294 (95%CI: 0.000744-0.153) False negative: 0.313 (95%CI: 0.161-0.500) **Algorithm VI** (1 uncertain SPECT) Start with CT. • Positive CT: 25/26 TP, 1/26 FP • Negative CT continue with MRI. ○ Negative MRI continue with SPECT. ▪ Negative SPECT: 13/13 TN ▪ Positive SPECT: 10/19 TP, 9/10 FP ○ Positive MRI: 8/8 TP Total false results: 10 False positive: 0.189 (95%CI: 0.0944-0.320) False negative: 0 (95%CI: 0-0.247) **Algorithm VII** N=65 Start with MRI. • Positive MRI: 27/27 TP • Negative MRI continue with SPECT. ○ Positive SPECT: 15/25 TP, 10/25 FP	

			Algorithm VI (1 uncertain SPECT) Start with CT. • Positive CT • Negative CT continues with MRI. ○ Negative MRI continues with SPECT. ▪ Negative SPECT ▪ Positive SPECT: ○ Positive MRI Algorithm VII Start with MRI. • Positive MRI: 27/27 TP • Negative MRI continues with SPECT. ○ Positive SPECT ○ Negative SPECT Algorithm VIII Start with CT. • Positive CT • Negative CT continues with SPECT. ○ Positive SPECT ○ Negative SPECT Comparator test: N/A Cut-off point(s): N/A			○ Negative SPECT: 13/13 TN Total false results: 10 False positive: 0.192 (95%CI: 0.0963-0.325) False negative: 0 (95%CI: 0-0.247) **Algorithm VIII** Start with CT. • Positive CT: 25/26 TP, 1/26 FN • Negative CT continue with SPECT. ○ Positive SPECT: 18/27 TP, 9/27 FP ○ Negative SPECT: 13/13 TP Total false results: 10 False positive: 0.189 (95%CI: 0.0944-0.320) False negative: 0 (95%CI: 0-0.247)	

Exclusietabel

Tabel Exclusie na het lezen van het volledige artikel

Algoritmen

Auteur en jaartal	Redenen van exclusie
Kalavrezos 1006	Bevatte enkel imaging data van vóór 2000
Wong 1996	Geen algoritme, gaat niet specifiek over mandibula invasie
Ord 1997	Geen algoritme
Schimming 2000	Geen algoritme
Imola 2001	Geen algoritme
Mukherji 2001	Geen algoritme
Werning 2001	Geen algoritme
Gu 2010	Geen algoritme
Arya 2013	Geen algoritme
Hakim 2014	Geen algoritme
Jamdade 2014	Geen algoritme
Kolk 2014	Geen algoritme
Li 2015	Geen algoritme
Li 2015	Artikel in de Chinese taal
Farrow 2016	Geen algoritme
Silva 2016	Geen algoritme
Qiao 2018	Geen algoritme
Nae 2019	Geen algoritme, maar diagnostische accuratesse van gecombineerde modaliteiten

Modaliteiten

Auteur en jaartal	Redenen van exclusie
Uribe 2013	Voegt qua studieselectie geen nieuwe informatie toe ten opzichte van Qiao 2018
Evangelista 2014	voegt qua studieselectie geen nieuwe informatie toe ten opzichte van Qiao 2018
Li 2014	voegt qua studieselectie geen nieuwe informatie toe ten opzichte van Qiao 2018, mist een aantal studies ten opzichte van Qiao 2018
Li 2014	Voegt qua studieselectie geen nieuwe informatie toe ten opzichte van Qiao 2018
Chun-Jie 2015	Artikel in de Chinese taal
Li 2015	Voegt qua studieselectie geen nieuwe informatie toe ten opzichte van Qiao 2018
Li 2015	Artikel in de Chinese taal
Sheng 2015	Conference abstract
Xiaonian 2017	Artikel in de Chinese taal
Brandao 2018	Mist veel studies over CT en MRI die wel in Qiao 2018 zitten

Zoekverantwoording

Algoritmen

Uitgangsvraag: Hoe dient botinvasie preoperatief van de mandibula bepaald te worden, door middel van diagnostische algoritmen	
Database(s): Ovid/Medline	Datum: 18-11-2019
Periode: niet van toepassing	Talen: niet van toepassing
Toelichting: In eerste instantie gezocht met de P en de I. Vanwege het geringe aantal referenties de keuze gemaakt om alleen met mandibular invasion te zoeken zonder de combinatie met mondholtecarcinoom en zonder een filtering op studiedesign toe te passen. Vr. groet, Ingeborg van Dusseldorp	

	Inclusief dubbele referenties	Ontdubbeld
Totaal	95	56

Zoekverantwoording

Ovid/Medline

1 exp Mandible/ or mandib*.ti,ab,kf. or exp Mandibular Neoplasms/ or jaw.ti,ab,kf. (130121)
2 exp Neoplasm Invasiveness/ or invasi*.ti,ab,kw. or involvement.ti,ab,kf. (942372)
3 1 and 2 (4930)
4 exp Algorithms/ or algorithm*.ti,ab,kf. or 'test batter*'.ti,ab,kf. or ('combin*' and (method* or examin*)).ti,ab,kf. (1135807)
5 3 and 4 (280)
6 exp "Sensitivity and Specificity"/ or (specificit* of screening or accura* or reference value* or false positive of false negative or predictive value or roc or likelyhood* or likelihood*).ti,ab,kf. (1393295)
7 5 and 6 (50)

Embase

No.	Query	Results
#9	#8 NOT 'conference abstract':it	**45**
#8	#1 AND #2 AND #3	**56**
#7	#5 AND #6	**4**
#6	('clinical trial'/exp OR 'randomization'/exp OR 'single blind procedure'/exp OR 'double blind procedure'/exp OR 'crossover procedure'/exp OR 'placebo'/exp OR 'prospective study'/exp OR rct:ab,ti OR random*:ab,ti OR 'single blind':ab,ti OR 'randomised controlled trial':ab,ti OR 'randomized controlled trial'/exp OR placebo*:ab,ti) NOT 'conference abstract':it	**2312047**
#5	#1 AND #2 AND #3 AND #4	**23**
#4	'mouth tumor'/exp OR 'oral cavity squamous cell carcinoma'/exp OR ((('buccal mucosa' OR intraoral OR 'oral' OR mouth OR tongue OR pharynx OR 'salivary gland*') NEAR/7 (tumor* OR tumour* OR carcinoma* OR cancer* OR neoplasm* OR squamous)):ti,ab)	**148660**
#3	'sensitivity and specificity'/exp OR 'screening'/exp OR 'reference value'/exp OR 'diagnostic error'/exp OR 'predictive value'/exp OR 'receiver operating characteristic'/exp OR specificit*:ab,ti OR screening:ab,ti OR accura*:ab,ti OR 'reference value*':ab,ti OR 'false positive':ab,ti OR 'false negative':ab,ti OR 'predictive value*':ab,ti OR roc:ab,ti OR likelyhood*:ab,ti OR likelihood*:ab,ti	**2955922**
#2	'algorithm'/exp OR 'algorithm':ti,ab OR 'test batter*':ti,ab OR ('combin*':ti,ab,kw AND (method*:ti,ab,kw OR examin*:ti,ab,kw))	**1432654**
#1	('mandible'/exp OR mandib*:ti,ab,kw OR 'lower jaw':ti,ab,kw OR 'jaw cancer'/exp OR 'jaw cancer':ti,ab,kw) AND ('tumor invasion'/exp OR invasi*:ti,ab OR involvement:ti,ab)	**4694**

Modaliteiten

Uitgangsvraag: Hoe dient botinvasie (preoperatief) van de mandibula bewaakt te worden? modaliteiten	
Database(s): Ovid/Medline, Embase	Datum: 20-11-2019
Periode: niet van toepassing	Talen: niet van toepassing
Toelichting: Alle sleutelartikelen werden gevonden. Vr. groet, Ingeborg van Dusseldorp	

	Inclusief dubbele referenties	Ontdubbeld
SR	31	22

Zoekverantwoording

Ovid/Medline

1 exp Diagnostic Imaging/ or exp radiography/ or 'spect'.ti,ab,kf. or 'single photon emission computed tomograph*'.ti,ab,kf. or 'single photon emission computer tomograph*'.ti,ab,kf. or 'dual energy ct'.ti,ab,kf. or 'dual energy computed tomography'.ti,ab,kf. or 'cone beam ct'.ti,ab,kf. or 'cone beam computed tomograph*'.ti,ab,kf. or 'cone beam computerized tomograph*'.ti,ab,kf. or 'cone-beam computed tomograph*'.ti,ab,kf. or 'spiral cone-beam computed tomograph*'.ti,ab,kf. or 'volume ct'.ti,ab,kf. or 'volume computed tomograph*'.ti,ab,kf. or 'volumetric ct'.ti,ab,kf. or 'volumetric computed tomograph*'.ti,ab,kf. or cbct.ti,ab,kf. or 'cat scan'.ti,ab,kf. or 'cat scanning'.ti,ab,kf. or 'computed tomographic scan'.ti,ab,kf. or 'computed tomograph*'.ti,ab,kf. or 'computed tomography scan'.ti,ab,kf. or 'computer assisted tomograph*'.ti,ab,kf. or 'computer tomograph*'.ti,ab,kf. or 'computerised axial tomograph*'.ti,ab,kf. or 'computerised tomograph*'.ti,ab,kf. or 'computerized axial tomograph*'.ti,ab,kf. or 'computerized tomograph*'.ti,ab,kf. or 'computerized tomography scan'.ti,ab,kf. or 'expert plus'.ti,ab,kf. or 'expert plus 1t'.ti,ab,kf. or 'mri'.ti,ab,kf. or 'nmr imaging'.ti,ab,kf. or 'imaging, magnetization transfer'.ti,ab,kf. or 'magnetic resonance imaging'.ti,ab,kf. or 'magnetic resonance tomograph*'.ti,ab,kf. or 'magnetization transfer imaging'.ti,ab,kf. or 'mr imaging'.ti,ab,kf. or 'nuclear magnetic resonance imaging'.ti,ab,kf. or 'orthopantography'.ti,ab,kf. or 'orthopantomograph*'.ti,ab,kf. or 'panoramic radiography'.ti,ab,kf. or 'pantomograph*'.ti,ab,kf. or 'radiography, panoramic'.ti,ab,kf. (2772799)
2 (exp Mandible/ or mandib*.ti,ab,kf. or exp Mandibular Neoplasms/ or jaw.ti,ab,kf.) and (exp Neoplasm Invasiveness/ or invasi*.ti,ab,kw. or involvement.ti,ab,kf.) (4931)
3 exp "Sensitivity and Specificity"/ or (specificit* of screening or accura* or reference value* or false positive of false negative or predictive value or roc or likelyhood* or likelihood*).ti,ab,kf. (1394731)
4 (meta-analysis/ or meta-analysis as topic/ or (meta adj analy$).tw. or ((systematic* or literature) adj2 review$1).tw. or (systematic adj overview$1).tw. or exp "Review Literature as Topic"/ or cochrane.ab. or cochrane.jw. or embase.ab. or medline.ab. or (psychlit or psyclit).ab. or (cinahl or cinhal).ab. or cancerlit.ab. or ((selection criteria or data extraction).ab. and "review"/)) not (Comment/ or Editorial/ or Letter/ or (animals/ not humans/)) (421944)
5 1 and 2 and 3 and 4 (13)

Embase

No.	Query	Results
#25	#2 AND #4 AND #16 AND #24	**18**
#24	#1 OR #20 OR #23	**2611701**
#23	'radiography'/exp	**1207427**
#22	#2 AND #4 AND #16 AND #21	**18**
#21	#1 OR #20	**1834816**
#20	'computer assisted diagnosis'/exp	**1098009**
#19	#17 AND #18	**6**
#18	#6 OR #7 OR #10 OR #12 OR #14	**61**
#17	#5 AND #16	**18**
#16	('meta analysis'/de OR cochrane:ab OR embase:ab OR psycinfo:ab OR cinahl:ab OR medline:ab OR ((systematic NEAR/1 (review OR overview)):ab,ti) OR ((meta NEAR/1 analy*):ab,ti) OR metaanalys*:ab,ti OR 'data extraction':ab OR cochrane:jt OR 'systematic review'/de) NOT (('animal experiment'/exp OR 'animal model'/exp OR 'nonhuman'/exp) NOT 'human'/exp)	**463146**
#15	#5 AND #14	**1**
#14	brandao AND magnetic AND mandibular AND invasion AND 2018	**1**
#13	#5 AND #12	**2**
#12	li AND computed AND tomography AND mandibular AND 2014	**47**

#11	#5 AND #10	**1**
#10	li AND magnetic AND mandibular AND 2014	**10**
#9	#5 AND #7	**3**
#8	#5 AND #6	**1**
#7	li AND emission AND mandibular AND 2015	**3**
#6	qiao AND performance AND mandibular AND 2018	**1**
#5	#3 AND #4	**277**
#4	'sensitivity and specificity'/exp OR 'screening'/exp OR 'reference value'/exp OR 'diagnostic error'/exp OR 'predictive value'/exp OR 'receiver operating characteristic'/exp OR specificit*:ab,ti OR screening:ab,ti OR accura*:ab,ti OR 'reference value*':ab,ti OR 'false positive':ab,ti OR 'false negative':ab,ti OR 'predictive value*':ab,ti OR roc:ab,ti OR likelyhood*:ab,ti OR likelihood*:ab,ti	**2959582**
#3	#1 AND #2	**1518**
#2	('mandible'/exp OR mandib*:ti,ab,kw OR 'lower jaw':ti,ab,kw OR 'jaw cancer'/exp OR 'jaw cancer':ti,ab,kw) AND ('tumor invasion'/exp OR invasi*:ti,ab OR involvement:ti,ab)	**4694**
#1	'single photon emission computed tomography'/exp OR 'spect':ti,ab OR 'single photon emission computed tomography':ti,ab OR 'single photon emission computer tomography':ti,ab OR 'dual energy computed tomography'/exp OR 'dual energy ct':ti,ab OR 'dual energy computed tomography':ti,ab OR 'cone beam computed tomography'/exp OR 'cone beam ct':ti,ab OR 'cone beam computed tomography':ti,ab OR 'cone beam computerized tomography':ti,ab OR 'cone-beam computed tomography':ti,ab OR 'spiral cone-beam computed tomography':ti,ab OR 'volume ct':ti,ab OR 'volume computed tomography':ti,ab OR 'volumetric ct':ti,ab OR 'volumetric computed tomography':ti,ab OR cbct:ti,ab OR 'computer assisted tomography'/exp OR 'cat scan':ti,ab OR 'cat scanning':ti,ab OR 'computed tomographic scan':ti,ab OR 'computed tomography':ti,ab OR 'computed tomography scan':ti,ab OR 'computer assisted tomography':ti,ab OR 'computer tomography':ti,ab OR 'computerised axial tomography':ti,ab OR 'computerised tomography':ti,ab OR 'computerized axial tomography':ti,ab OR 'computerized tomography':ti,ab OR 'computerized tomography scan':ti,ab OR 'nuclear magnetic resonance imaging'/exp OR 'expert plus':ti,ab OR 'expert plus 1t':ti,ab OR 'mri':ti,ab OR 'nmr imaging':ti,ab OR 'imaging, magnetization transfer':ti,ab OR 'magnetic resonance imaging':ti,ab OR 'magnetic resonance tomography':ti,ab OR 'magnetization transfer imaging':ti,ab OR 'mr imaging':ti,ab OR 'nuclear magnetic resonance imaging':ti,ab OR 'panoramic radiography'/exp OR 'orthopantography':ti,ab OR 'orthopantomography':ti,ab OR 'panoramic radiography':ti,ab OR 'pantomography':ti,ab OR 'radiography, panoramic':ti,ab	**1793177**

Module 2.1 HPV-statusbepaling

Uitgangsvraag

Hoe moet de HPV-status bepaald worden?

De uitgangsvraag omvat de volgende deelvragen:

1. Hoe moet de HPV-status op histologisch materiaal bepaald worden bij patiënten met een gediagnosticeerd orofarynx carcinoom?
2. Hoe moet de HPV-status bepaald worden bij patiënten met een lymfekliermetastase in de hals van een onbekende primaire tumor?

Inleiding

De Humaan Papillomavirus (HPV) status van orofarynx carcinomen kan bepaald worden met verschillende testen. Op dit moment is het onduidelijk welke op histopathologie gebaseerde teststrategie de beste diagnostische accuratesse voor het bepalen van de HPV-status van gediagnosticeerde orofarynx carcinomen heeft. In sommige omstandigheden is histopathologisch materiaal niet beschikbaar, bijvoorbeeld wanneer de primaire tumor onbekend is. De HPV-status zou dan wellicht op basis van cytologische tests kunnen worden vastgesteld op cytologisch materiaal dat met een lymfklierpunctaat is verkregen bij patiënten met een onbekende primaire tumor en een gediagnosticeerde halsmetastase. Echter is het op dit moment nog onduidelijk wat de diagnostische accuratesse van HPV-testen op cytologisch materiaal van een positieve lymfeklier uit de hals is.

Search and select

A systematic review of the literature was performed to answer the following questions:

PICO 1

What is the diagnostic accuracy of diagnostic test algorithms to determine the HPV-status on histological material in patients with an oropharyngeal carcinoma?

P: patients with an oropharyngeal carcinoma;
I: diagnostic strategies/algorithms to determine the HPV-status based on histopathologic tests;
C: diagnostic strategies/algorithms compared;
R: a test to detect HPV-DNA and/or E6/E7 mRNA;
O: sensitivity, specificity, positive predictive value, negative predictive value.

PICO 2

What is the diagnostic accuracy of tests on cytologic material to determine the HPV-status in patients with a carcinoma of unknown primary?

P: patients with a carcinoma of unknown primary and a positive neck node;
I: diagnostic tests to determine the HPV-status based on cytologic material;
C: comparison of tests on cytologic material;
R: a test to detect HPV-DNA and/or E6/E7 mRNA;
O: sensitivity, specificity, positive predictive value, negative predictive value.

<u>Relevant outcome measures</u>

The guideline development group considered sensitivity and negative predictive value as a critical outcome measures for decision making; and specificity and positive predictive value as an important outcome measures for decision making.

A priori, the working group did not define the outcome measures listed above but used the definitions used in the studies.

Search and select (Methods)

PICO 1

The databases Medline (via OVID) and Embase (via Embase.com) were searched with relevant search terms from 2016 until the 10th of April 2020 for PICO 1 (histology). The time limiter was chosen because the guideline "*Human Papillomavirus Testing in Head and Neck Cancers*" from the College of American Pathologists (CAP) had their latest searched performed in 2016 (Lewis, 2018). The detailed search strategy is depicted under the tab Methods. The systematic literature search resulted in 251 hits. Studies were selected based on the following criteria: patients had an oropharyngeal carcinoma, diagnostic strategies or algorithms were used to determine the HPV-status with histopathological tests, the reference test was a test that detected HPV DNA and/or mRNA, and at least one of the outcomes of interest was reported or it could be calculated manually from the presented data. Conference abstracts and non-systematic reviews were excluded. A total of 6 studies were initially selected based on title and abstract screening. After reading the full text, 5 studies were excluded (see the table with reasons for exclusion under the tab Evidence tables), and 1 systematic review (which included 24 primary studies) was included.

PICO 2

The databases Medline (via OVID) and Embase (via Embase.com) were searched with relevant search terms until the 27th of February 2020 for PICO 2 (cytology). The detailed search strategy is depicted under the tab Methods. The systematic literature search resulted in 448 hits. Studies were selected based on the following criteria: patients had a (suspected) neck metastasis, patients had (suspected) primary head and neck squamous cell carcinoma at the time material was taken for cytologic tests, the reference test was a test that detected HPV DNA and/or mRNA, and at least one of the outcomes of interest were reported or it could be calculated manually from the presented data. Conference abstracts and non-systematic reviews were excluded. A total of 74 studies were initially selected based on title and abstract screening. After reading the full text, 62 studies were excluded (see the table with reasons for exclusion under the tab Evidence tables) and 12 studies were included.

Results

One systematic review (including 24 primary studies) was included in the analysis of the literature for the histopathology-based diagnostic algorithms. Important study characteristics and results are summarized in the evidence tables. The assessment of the risk of bias is summarized in the risk of bias tables (under the tab Evidence tables).

Thirteen studies were included in the analysis of the literature for the cytology-based tests. Important study characteristics and results are summarized in the evidence tables. The assessment of the risk of bias is summarized in the risk of bias tables. Data regarding the classification of cases by the index test compared to the reference test were extracted from each of the included studies. Diagnostic accuracy parameters and/or 95% confidence intervals were calculated based on the extracted data when not reported in the original study.

Summary of literature

Description of studies

Diagnostic algorithms on histological material to determine the HPV-status in patients with a confirmed oropharyngeal carcinoma (PICO 1)

Prigge (2017) conducted a systematic review where the diagnostic accuracy of a strategy was assessed where $p16^{INK4a}$ immunohistochemistry and an HPV DNA PCR was combined for HPV-testing in oropharyngeal squamous cell carcinomas. MEDLINE was searched through

PubMed on the 8th of January 2016. Studies were included when persons in the sample were diagnosed with oropharyngeal squamous cell carcinoma, $p16^{INK4a}$ was the index test, a reference test was used that detected E6 and/or E7 mRNA and when the study design was prospective or retrospective. Studies were excluded when the authors of the original studies did not respond to inquiries about the presented data, when the sample size was smaller than 10, and when the study did not report primary data. Eleven of the included studies reported the use of $p16^{INK4a}$ and HPV DNA PCR as a combined test for diagnostic test accuracy evaluation against a reference standard that detected E6 and/or E7 mRNA. The eleven studies comprised of a sample of 509 persons (as reported in the study characteristics table in the systematic review). A case was defined as positive when both the p16INK4a and the HPV DNA PCR returned positive results. When one of both or both tests returned a negative result, the case was defined as negative. Studies were assessed with the QUADAS-2 tool for risk of bias and applicability. Most of the 11 studies scored 'unclear' in the patient selection domain, where it was mostly unclear whether studies avoided inappropriate exclusions. The applicability regarding the patient selection was generally judged as having moderate concerns. The cut-off for defining positive/negative cases by using $p16^{INK4a}$ varied among the included studies and was tested on whole tissue section FFPE material (with one exception, where tissue microarray was used on FFPE material). Five studies used the G175-405 antibody clone for $p16^{INK4a}$, while the other six studies used E6H4. Variation in procedures was also observed in the HPV DNA PCR methods. Here, GP5+/6+ (reverse line blot genotyping or bead-based genotyping), HPV 16 primers, HPV 16 E6/E7 primers, MY09/MY11/HMB01 (dot blot hybridization genotyping), BSGP5+/6+ (bead-based genotyping) were the described methods. Reference tests also had variation in their procedures. Transcript types used were E6, E6*I, E7, or combinations thereof. Reference tests also showed variation in the detected HPV types. HPV 16 was sought for detection in all studies, however some studies added HPV 18, 26, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 67, 68b, 70, 73, and/or 82 for detection as well. Four of the eleven studies used whole tissue section FFPE material for the reference test, while the remaining seven studies used fresh frozen material.

Tests on cytological material in patients with positive neck nodes and CUP (PICO 2)
Baldassari (2015) used a cobas HPV assay on fine needle aspirates and compared it to a combined test of p16 and HPV ISH on paraffin-embedded formalin-fixed surgical tissue of the primary and/or metastatic tumor. Specimens were collected prospectively, but it remained unclear whether this was consecutively. Inclusion and exclusion criteria were not described. Air-dried and alcohol fixed smears were prepared and stained. A slide was prepared after centrifugation and the cobas HPV assay was performed according to the manufacturer's protocol. Fourteen HPV types were targeted for amplification (HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68). Both p16 IHC and HPV ISH were performed on deparaffinized 5-micrometer sections. For p16, sections were incubated with a mouse monoclonal antibody (E6H4). Sections for HPV ISH were incubated with the INFORM HPV III Family 16 probe, detecting HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66. The cut-off points for a positive/negative case in the cobas HPV assay, p16, and HPV ISH were unclear. Thirty-seven participants were recruited and forty-two fine needle aspirates were taken. Participants had a mean age of 60.4 (SD: 11.6) years. Seventeen participants had a history of head and neck squamous cell carcinomas. Fine needle sample sites were the neck mass (n=36), the mediastinal lymph node (n=5), or a left parapharyngeal mass (n=1).

Begum (2007) searched a database and selected cases when the processing of the initial fine needle aspirate included the preparation of a cellblock by spinning the cell block in a cellular pellet. The index test was p16 on cell blocks, where 5-micron sections were deparaffinized.

Sections were then incubated with a mouse monoclonal antibody. Observing any staining in the squamous cells was considered to be positive for HPV. Material from fine needle aspirates were also tested with HPV16 ISH (considered as a reference test). HPV16 ISH was performed on cell blocks for signal amplification and on resections of the primary tumor. Signals visualized as dots in nuclei of the squamous cells were considered positive for HPV. Nineteen participants with oropharyngeal tumors and ten participants with an unknown primary were selected. No other patient characteristics were described.

Bishop (2012) consecutively recruited participants to assess the accuracy of the Hybrid Capture 2 assay (HC2), although inclusion and exclusion criteria were not reported. Metastatic tumors were aspirated using a 12 gauge needle and multiple passes. HC2 detected HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 68. Specimen DNA was denatured into single stranded DNA. RNA/DNA hybrids were then immobilized onto a microplate surface. Light is emitted and measured in relative light units. A case with ≥ 3 RLU/CO was considered positive for HPV, 0.85-3 RLU/CO was considered equivocal, and < 0.85 RLU/CO was considered negative. HPV ISH was used as the reference test. Hybridization was performed using the HPV III Family 16 probe (HPV16, 18, 33, 35, 45, 51, 52, 56, 66) on 5-micron section from the tissue microarray was assessed. HPV ISH was considered positive when signals localized to tumor cell nuclei. Participant recruitment resulted in 24 participants (27 cytologic preparations), of which 12 had a lymph node sample site. From these 12 participants, the tumor site was the skin (n=2), larynx (n=2, floor of mouth (n=1), tongue (n=1), base of tongue (n=1), tonsil (n=4), and unknown (n-1).

Buonocore (2019) recruited 25 participants consecutively (n=24 were positive for HPV by HPV ISH, n=1 was non-contributory). Participants were included when they had previous or unknown oropharyngeal head and neck squamous cell carcinoma (or this was determined at the time of the procedure) and an unknown p16 status. Exclusion criteria were not reported. Fine needle aspirates were performed with a 25 gauge needle. Diff-Quik-stained and ethanol fixed smears were prepared. Passes were allowed to clot before fixed in formalin. From this material both a CytoLyt-fixed and a formalin-fixed cell block were made. Both the CytoLyt-fixed and formalin-fixed cell blocks were tested with p16 (indextest) against HPV ISH (reference test). Mouse monoclonal antibodies (E6H4) were used for p16. A cut-off of ÷70% staining in nuclei and cytoplasm was used to define a HPV-positive case. It was unclear on which material the HPV ISH was performed and it was unclear how an HPV-positive case was defined. HPV ISH targeted HPV 16, 18, 26, 31, 33, 35, 39, 41, 45, 51, 52, 53, 56, 58, 59, 66, 68, 73, 82, and E6/E7 mRNA. Participants (22 male; 3 female) had a mean age of 60 (range: 47 to 76) years and a variety in smoking history (never smoked: n=14, smoking history: n=11 (range: 0.5-40 pack years)). Similarly, a variety of alcohol use was observed: never (n=1), abstinent (n=1), occasional (n=2), social (n=15), daily (n=3), heavy (n=3).

Chute (2014) recruited 95 participants (resulting in 96 fine needle aspirates) prospectively. Participants were eligible when a fine needle aspirate from a head and neck-site was interpreted as being a squamous cell carcinoma, atypical, or suspicious for squamous cell carcinoma. Exclusion criteria were not reported. A cell block was made in an automated system according to its manufacturer's directions. However, methylene blue was replaced by eosin. HC2 and CISH were performed on cytological material. HC2 was performed according to the manufacturer's instructions, targeting HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68. A RLU/CO ≥ 1 was defined as a HPV-positive case. For, CISH, the HPV III Family 16 probe was used (HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58). A discrete blue dot-like staining in the tumor nuclei was defines as an HPV-positive case. CISH combined with p16 was performed on a surgical biopsy. An HPV-positive case for p16 was defined as > 75%

strong and diffuse cytoplasmic and nuclear staining. For the CISH and p16 combined test, an HPV-positive case was defined as having a positive test result from both CISH and p16. Formalin-fixed paraffin-embedded tissue of the excised primary tumor or neck metastasis were used for testing in the CISH and p16 combined test. Participants (72 male; 21 female) had a median age of 60 (range: 17-93) years. The primary tumor location was oropharyngeal (n=32), non-oropharyngeal (n=32), non-head and neck (n=18), or unknown (n=13).

Hou (2016) searched a database to select cases with metastatic head and neck squamous cell carcinomas in cervical lymph nodes, diagnosed by a fine needle aspirate. HPV ISH and p16 had to be performed on fine needle aspirate material to be selected from the database. No exclusion criteria were reported. Both p16 (index test) and HPV ISH (reference test) were performed on cytologic material. Fine needle aspirates were centrifuged and the specimen was clot dried. The specimen was then placed in a CellSafe mesh capsule and fixated in formalin (10% neutral-buffered). For p16, monoclonal antibodies (E6H4) were used. An HPV-positive case by p16 was defined as ≥ 70% diffuse or strong nuclear or cytoplasmic staining. HPV ISH was performed according to the manufacturer's protocol on 4-micrometer sections of the cell block. HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 68, and Y1443 were targeted. Presence of staining in the nuclei defined an HPV-positive case for HPV ISH. Participants (80 male; 7 female) had a mean age of 59 (range: 38 to 86) years. The primary site of the tumor was the base of the tongue (n=32), tonsil (n=19), other oropharyngeal sites (not specified, n=4), non-oropharyngeal (not specified, n=25), or unknown (n=7).

Jalaly (2015) searched a database to select cases that had a metastatic cervical lymph node (proven by fine needle aspirates) with a corresponding surgical specimen (either biopsy or resection). P16 and HPV ISH were performed on the cell blocks created from the fine needle aspirate. Fine needle aspirate material was centrifuged for 2 minutes. The specimen clot was allowed to dry on tissue paper. Thereafter, it was placed in a CellSafe capsule and fixed in formalin (10% neutral-buffered). For p16, a monoclonal antibody (E6H4) was used and an HPV-positive case for p16 was defined as >70% nuclear and cytoplasmic staining of the tumor cells. HPV ISH detected E6/E7 mRNA and was performed according to the manufacturer's instruction and 4-millimeter formalin-fixed paraffin-embedded cell block sections were prepared. Red punctate dots in the nucleus and/or cytoplasm signals higher than the signal on a DapB-negative control slide was defined as a HPV-positive case. Forty-eight participants were recruited (44 male; 4 female). The fine needle aspirate sample site was the neck (n=41), subcarinal (n=2), mediastinal (n=1), submandibular (n=2), chest wall (n=1), or supraclavicular (n=1). The specimen was either resected (n=32) or a biopsy was made (n=16). The tumor site of the primary tumor was at the base of the tongue (n=14), tonsil (n=15), other oropharyngeal (not specified, n=8), oral cavity (n=6), larynx (n=1), maxilla (n=1), or unknown (n=3).

Janapureddy (2010) searched a database to select participants that had a cell block cytologic diagnosis of metastatic squamous cell carcinoma in a cervical lymph node. Participants with inadequate cell block material were excluded. Cytologic material was tested with $p16^{INK4a}$ and ProExC as index tests and compared to HPV ISH on cell block sections. Material from fine needle aspirates were fixed in formalin (10% neutral-buffered). After centrifugation the supernatant was discarded and the resulting content was assessed. Cell block tissue was created (5-micrometer) from the formalin-fixed paraffin-embedded tissue. Incubation was performed with E6H4 monoclonal $p16^{INK4a}$ at room temperature. An HPV-positive case for p16 was defined as the presence of nuclear and cytoplasmic staining. ProExC also had in incubation period at room temperature. Presence of nuclear staining defined an HPV-positive case for ProExC. HPV ISH detected HPV 16, 18, 31, 33, and 51. Cell block tissue was

deparaffinized and rehydrated. Slides were air dried, sections were denatured and hybridized. HPV-positive cases by HPV ISH were defined as the presence of punctate or dot-like nuclear staining. Participants (36 male; 4 female) had a mean age of 58.2 (reange 25 to 87) years. The primary tumor site was oropharyngeal (not specified, n=11), nasopharyngeal (n=2), other (not specified, n=5), or was not determined (n=9).

Sivars (2017) prospectively obtained fine needle aspirate material. Participants were recruited when they were suspected of head and neck carcinoma or had a neck mass suspicious for metastasis, and when there was not enough material left for HPV testing after cytological diagnosis. The HPV-status was tested on cytological material from fine needle aspirates and/or formalin-fixed paraffin-embedded material (either resection or biopsy). For cytologic testing, DNA was extracted from fine needle aspirates. The DNA multiplex assay was performed by using GP5+/GP6+ primers and additionally E6 (HPV 16, 33) was amplified. DNA detection was performed on a bead-based multiplex using mean fluorescent intensity. An HPV-positive case for the multiplex assay was defined as a mean fluorescent intensity above the background * 1.5 + 15. Furthermore, a Real-Time PCR was used in the clinic to detect seven common high-risk HPV genotypes. Is was unclear how a positive/negative case was defined vor the Real-Time PCR in the clinic. Sixty-six patients (35 male; 31 female) participated. Fine needle aspirate sample sites were the primary tumor (n=2) or neck masses (n=64). The mean age was 61 year for persons with an oropharyngeal tumor (n=20), 71.5 years for persons with other malignancies (n=17), and 53 years for persons with benign conditions (n=29).

Smith (2014) recruited participants prospectively when a cervical lymph node was swollen to one centimeter or larger (exclusion criteria were not reported). A modified HC2 HPV assay detected HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68 in cytologic material. Fine needle biopsies of cervical metastases were performed with a 25 gauge needle. The aspirate was placed on a slide and was air-dried and stained. A final pass was made with a fresh needle in to the lymph node, which was stored until used for HC2. DNA was denatured and incubated for 45 minutes. Samples were applied to hybrid-specific antibodiy coated microplate wells. Signal amplification was performed with Detection Reagent II and light emission was used to detect HPV DNA. An HPV-positive case for HC2 was defined as ≥ 2.5 RLO/CO, an equivocal case as 0.85 to 2.5 RLU/CO, and a negative case as < 0.85 RLU/CO. HPV ISH was performed on tissue specimen from resected participants. Five-micron formalin-fixed paraffin-embedded tissue sections from either tumors or biopsies were used for HPV ISH. The HPV III Family 16 probe set was used to detect HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, and 66). Punctate signals in the nuclei defined a positive HPV case for HPV ISH. The mean age and sex distribution were not reported for the participants, however 25 persons were recruited. The tumor location was on the palatine tonsil (n=6), base of the tongue (n=8), hypopharynx (n=1), skin of the auricle (n=1), or unknown (n=2).

Takes (2016) searched a database for cases where formalin-fixed paraffin-embedded histological material was available and was tested positive or negative for both HPV and p16. Cases were excluded when there was a secondary tumor in the head and neck region, when there was not enough cytological material, or when there was previous exposure to radiotherapy. A HPV PCR was performed on fine needle aspirates material scraped from archival slides. The DNA was purified, diluted and stored until tested by HPV PCR. A broad-spectrum DNA amplification was performed in the HPV PCR. Probes were used in a micro titer hybridization assay. Cases positive in the micro titer hybridization assay were tested with line-specific probes (LiPA25) for detection of HPV 1, 16, 18, 31, 33, 34, 35, 39, 40, 42, 43, 44, 45, 51, 52, 53, 54, 56, 58, 59, 66, 68/73, 70, and 74. LiPA strips were visually

inspected and interpreted. Interpretation was performed following the standardized reference guide. It was unclear how an HPV-positive case was defined. Both the same HPV PCR and p16 were performed. The HPV PCR was performed on DNA isolated from formalin-fixed paraffin-embedded tissue sections of 4-micrometer. The p16 procedures were not reported. It was unclear, for both the HPV PCR and p16, how a HPV-positive case was defined. Participants (33 male; 14 female) had a mean age of 58 (range: 28.9 to 77.2) years. Their N-stage was N0 (n=6), N1 (n=7), or N2 (n=38). The primary tumor site was on the tonsils (n=19), the base of the tongue (n=12), other oropharyngeal (not specified, n=10), or unknown (n=6). Formalin-fixed paraffin-embedded material originated from the tonsils (n=21), base of the tongue (n=9), neck dissection (n=7), other oropharyngeal (not specified, n=10).

Xu (2016) searched a database and selected cases with cervical metastatic squamous cell carcinomas diagnosed with fine needle aspirates. For the selection, cases had to have p16 performed on both the cytological material and corresponding surgical material. Cases were excluded when tumors originated from outside the head and neck region. Cytologic material was prepared from fine needle aspirates for cell blocks and ThinPrep in CytoLyt solution. The p16 index test was performed on cell bock, smear or ThinPrep with pre-defined thresholds of 1%, 5%, 10%, 15%, and 70% nuclear and cytoplasmic staining to define an HPV-positive case. HPV CISH was performed on cytologic material using high risk HPV probes for detecting HPV 16, 18, 31, 33, and 51. An HPV-positive case for CISH was defined as discrete dot-like stippled nuclear labelling. How HPV-negative and equivocal cases were defined was not reported.

Results

Diagnostic algorithms on histological material to determine the HPV-status in patients with a confirmed oropharyngeal carcinoma (PICO 1)

Sensitivity

Prigge (2017) calculated a summary estimate for the sensitivity of $p16^{INK4a}$ combined with HPV DNA PCR from 11 studies (n=509). There was variation in underlying procedures (for example cut-offs for p16 positivity, materials, reference standards). Prigge (2017) found a pooled sensitivity 0.93 (95%CI: 0.87 to 0.97). Statistical heterogeneity (I^2) was 23.39%.

Specificity

Prigge (2017) pooled the specificity of a $p16^{INK4a}$ and HPV DNA PCR combined test from 11 studies (n=509). Underlying procedures showed variation in procedures (for example cut-offs for p16 positivity, materials, reference standards). A summary estimate was calculated and Prigge (2017) reported a specificity of 0.96 (0.89 to 1.00). Statistical heterogeneity (I^2) was 68.4%.

Positive predictive value

Positive predictive values were not reported.

Negative predictive value

Negative predictive values were not reported.

Tests on cytological material in patients with positive neck nodes and CUP (PICO 2)

Sensitivity

A variety of index tests on cytologic material were found. Tests were evaluated using several different methods and procedures: p16 (on CUP only, on oropharyngeal carcinoma only, on CytoLyt or formalin fixed cytologic material, at several thresholds, reference test on

cytological or histological material), HC2 (various reference tests, reference test on cytological or histological material), PCR or RT-PCR (various reference tests, reference test on cytological or histological material), ProExC, HPV CISH, and cobas 4800. Data could not be pooled due to heterogeneity in the study procedures. Sensitivity ranged from 0.00 to 1.00. An overview of the sensitivities (including 95% confidence intervals) are found in Figure 2.1. Sensitivity and/or 95% confidence intervals were calculated when not reported in the original study. Cases were excluded in most analyses (described in the evidence table, for example due to invalid or equivocal test results, no reference test performed, or inadequate specimens).

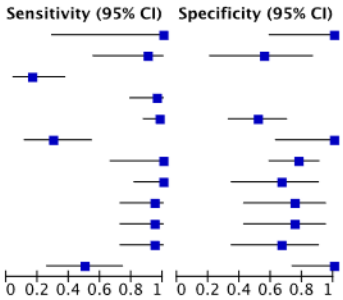
Specificity
A variety of index tests on cytologic material were found. A variety of index tests on cytological material were found, identical to as described under the sensitivity results. Data could not be pooled due to heterogeneity in the study procedures and tests. Specificity and/or 95% confidence intervals were calculated when not reported in the original study. Specificity ranged from 0.16 to 1.00. An overview of the specificities (including 95% confidence intervals) are found in Figure 2.1 Cases were excluded in most analyses (described in the evidence table, for example due to invalid or equivocal test results, no reference test performed, or inadequate specimens).

Positive and negative predictive value
Most studies did not report the positive and/or negative predictive values. When not reported, positive and negative predictive values were calculated from the extracted data. Positive and negative predictive values ranged from 0.00 to 1.00. Cases were excluded in most analyses (described in the evidence table, for example due to invalid or equivocal test results, no reference test performed, or inadequate specimens).

Figure 2.1 An overview of the sensitivity and specificity of the tests reported in the included references by index test, reference test and testing material. Category titles show the used index test (type of material) versus reference test (type of material)

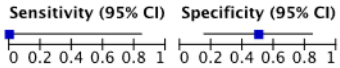
p16 (cytology) vs HPV ISH (cytology)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Begum 2007 (CUP only)	3	0	0	7	1.00 [0.29, 1.00]	1.00 [0.59, 1.00]
Begum 2007 (oropharynx only)	9	4	1	5	0.90 [0.55, 1.00]	0.56 [0.21, 0.86]
Buonocore 2019 (CytoLyt fixed)	4	0	20	0	0.17 [0.05, 0.37]	Not estimable
Buonocore 2019 (Formalin fixed)	23	0	1	0	0.96 [0.79, 1.00]	Not estimable
Hou 2016	41	15	1	16	0.98 [0.87, 1.00]	0.52 [0.33, 0.70]
Jalaly 2015	6	0	14	8	0.30 [0.12, 0.54]	1.00 [0.63, 1.00]
Jannapureddy 2010	9	7	0	24	1.00 [0.66, 1.00]	0.77 [0.59, 0.90]
Xu 2016 (1% threshold for p16)	18	4	0	8	1.00 [0.81, 1.00]	0.67 [0.35, 0.90]
Xu 2016 (10% threshold for p16)	17	3	1	9	0.94 [0.73, 1.00]	0.75 [0.43, 0.95]
Xu 2016 (15% threshold for p16)	17	3	1	9	0.94 [0.73, 1.00]	0.75 [0.43, 0.95]
Xu 2016 (5% threshold for p16)	17	4	1	8	0.94 [0.73, 1.00]	0.67 [0.35, 0.90]
Xu 2016 (70% threshold for p16)	9	0	9	12	0.50 [0.26, 0.74]	1.00 [0.74, 1.00]



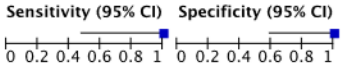
p16 (cytology) vs HPV ISH (histology)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Begum 2007 (oropharynx only)	0	4	2	4	0.00 [0.00, 0.84]	0.50 [0.16, 0.84]



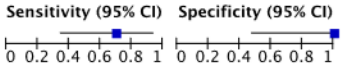
HC2 (cytology) vs HPV ISH (cytology)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bishop 2012	5	0	0	7	1.00 [0.48, 1.00]	1.00 [0.59, 1.00]



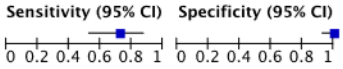
HC2 (cytology) vs HPV CISH (histology)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Smith 2014	7	0	3	5	0.70 [0.35, 0.93]	1.00 [0.48, 1.00]



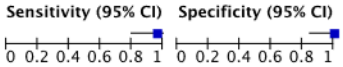
HC2 (cytology) vs HPV CISH + p16 (histology)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Chute 2014	21	0	8	48	0.72 [0.53, 0.87]	1.00 [0.93, 1.00]



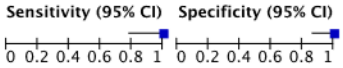
PCR (cytology) vs PCR + p16 (histology)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Takes 2016	24	0	1	22	0.96 [0.80, 1.00]	1.00 [0.85, 1.00]



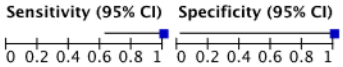
RT-PCR (cytology) vs Multiplex assay (cytology)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Sivars 2007	15	0	0	25	1.00 [0.78, 1.00]	1.00 [0.86, 1.00]



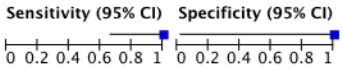
RT-PCR (cytology) vs Multiplex assay (histology)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Sivars 2007	8	0	0	1	1.00 [0.63, 1.00]	1.00 [0.03, 1.00]



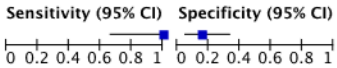
Multiplex assay (cytology) vs Multiplex assay (histology)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Sivars 2007	9	0	0	1	1.00 [0.66, 1.00]	1.00 [0.03, 1.00]



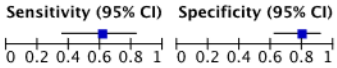
ProExC (cytology) HPV ISH (cytology)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Jannapureddy 2010	9	26	0	5	1.00 [0.66, 1.00]	0.16 [0.05, 0.34]



HPV CISH (cytology) vs HPV CISH + p16 (histology)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Chute 2014	11	7	7	27	0.61 [0.36, 0.83]	0.79 [0.62, 0.91]



Cobas 4800 (cytology) vs HPV ISH + p16 (histology)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Baldassari 2015	15	16	0	3	1.00 [0.78, 1.00]	0.16 [0.03, 0.40]

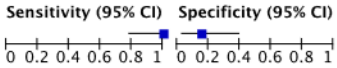


Table 2.2 Overview of the positive and negative predictive values of the tests on cytologic material by index test, reference test and testing material

Author, year (condition)	TP	FP	FN	TN	PPV	NPV
P16 (cytology) versusHPV ISH (cytology)						
Begum 2007 (CUP only)	3	0	0	7	1.00*	1.00*
Begum 2007 (oropharynx only)	9	4	1	5	0.69	0.833
Buonocore 2019 (CytoLyt fixed)	4	0	20	0	1.00*	0.00*
Buonocore 2019 (Formalin fixed)	23	0	1	0	1.00*	0.00*
Hou 2016	41	15	1	16	0.73	0.94
Jalaly 2015	6	0	14	8	1.00*	0.36
Jannapureddy 2010	9	7	0	24	0.56	1.00*
Xu 2016 (1% threshold for p16)	18	4	0	8	0.82	1.00*
Xu 2016 (10% threshold for p16)	17	3	1	9	0.85	0.90
Xu 2016 (15% threshold for p16)	17	3	1	9	0.85	0.90
Xu 2016 (5% threshold for p16)	17	4	1	8	0.81	0.89
Xu 2016 (70% threshold for p16)	9	0	9	12	1.00*	0.57
P16 (cytology) versus HPV ISH (histology)						
Begum 2007 (oropharynx only)	0	4	2	4	0.00*	0.67
HC2 (cytology) versus HPV ISH (cytology)						
Bishop 2012	5	0	0	7	1.00*	1.00*
HC2 (cytology) versus HPV CISH (histology)						
Smith 2014	7	0	3	5	1.00*	0.63
HC2 (cytology) versus HPV CISH = p16 (histology)						
Chute 2014	21	0	8	48	1.00*	0.86
PCR (cytology) versus PCR + p16 (histology)						
Takes 2016	24	0	1	22	1.00*	0.96
RT-PCR (cytology) versus Multiplex assay (cytology)						
Sivars 2007	15	0	0	25	1.00*	1.00*
RT-PCR (cytology) versus Multiplex assay (histology)						
Sivars 2007	8	0	0	1	1.00*	1.00*
Multiplex assay (cytology) versus Multiplex assay (histology)						
Sivars 2007	9	0	0	1	1.00*	1.00*
ProExC (cytology) versus HPV ISH (cytology)						
Jannapureddy 2010	9	26	0	5	0.26	1.00*
Cobas 4800 (cytology) versus HPV ISH + p16 (histology)						
Baldassari 2015	15	16	0	3	0.48	1.00*
HPV ISH (cytology) versus HPV ISH (histology)						
Begum 2007 (oropharynx only)	1	0	1	8	1.00*	0.89
***Calculation contained a cell value of zero**						

Level of evidence of the literature

Diagnostic algorithms on histological material to determine the HPV-status in patients with a confirmed oropharyngeal carcinoma (PICO 1)

The level of evidence regarding the outcome measure sensitivity (for p16^{INK4a} + HPV DNA PCR) was downgraded by 3 levels because of study limitations (1 level for risk of bias: most of the relevant studies were appraised by the authors as unclear regarding inappropriate

exclusions); applicability (1 level for bias due to indirectness: most of the relevant studies were appraised by the authors as having moderate applicability concerns regarding patient selection); number of included patients (1 level for imprecision: n=509 according to the general characteristics table); publication bias was not assessed.

The level of evidence regarding the outcome measure specificity (for $p16^{INK4a}$ + HPV DNA PCR) was downgraded by 3 levels because of study limitations (1 level for risk of bias: most of the relevant studies were appraised by the authors as unclear regarding inappropriate exclusions); applicability (1 level for bias due to indirectness: most of the relevant studies were appraised by the authors as having moderate applicability concerns regarding patient selection); number of included patients (1 level for imprecision: n=509 according to the general characteristics table); publication bias was not assessed.

The positive and negative predictive value was not reported and therefore GRADE was not performed.

Tests on cytological material in patients with positive neck nodes and CUP (PICO 2)
The level of evidence regarding the outcome measure sensitivity was downgraded by 4 levels because of study limitations (1 levels for risk of bias: 27.1% judgements in the QUADAS-2 were high risk of bias (predominantly in patient selection and flow/timing), 41.7% judgements were unclear risk (predominantly in index test and reference standard)); not downgraded for conflicting results (inconsistency: observed heterogeneity might possibly be explained by differences in procedures and thresholds); applicability (1 level for bias due to indirectness: mostly unclear whether the included patients matched the review question); number of included patients (2 levels for imprecision: very low number of participants per comparison); publication bias was not assessed.

The level of evidence regarding the outcome measure specificity was downgraded by 4 levels because of study limitations (1 levels for risk of bias: 27.1% judgements in the QUADAS-2 were high risk of bias (predominantly in patient selection and flow/timing), 41.7% judgements were unclear risk (predominantly in index test and reference standard)); not downgraded for conflicting results (inconsistency: observed heterogeneity might possibly be explained by differences in procedures and thresholds); applicability (1 level for bias due to indirectness: mostly unclear whether the included patients matched the review question); number of included patients (2 levels for imprecision: very low number of participants per comparison); publication bias was not assessed.

The level of evidence regarding the outcome measure positive predictive value was downgraded by 4 levels because of study limitations (1 levels for risk of bias: 27.1% judgements in the QUADAS-2 were high risk of bias (predominantly in patient selection and flow/timing), 41.7% judgements were unclear risk (predominantly in index test and reference standard)); not downgraded for conflicting results (inconsistency: observed heterogeneity might possibly be explained by differences in procedures and thresholds); applicability (1 level for bias due to indirectness: mostly unclear whether the included patients matched the review question); number of included patients (2 levels for imprecision: very low number of participants per comparison); publication bias was not assessed.

The level of evidence regarding the outcome measure negative predictive value was downgraded by 4 levels because of study limitations (1 levels for risk of bias: 27.1% judgements in the QUADAS-2 were high risk of bias (predominantly in patient selection and

flow/timing), 41.7% judgements were unclear risk (predominantly in index test and reference standard)); not downgraded for conflicting results (inconsistency: observed heterogeneity might possibly be explained by differences in procedures and thresholds); applicability (1 level for bias due to indirectness: mostly unclear whether the included patients matched the review question); number of included patients (2 levels for imprecision: very low number of participants per comparison); publication bias was not assessed.

Conclusions

Very low GRADE	There is a very low confidence in the sensitivity (0.93, 95%CI: 0.87 to 0.97) and specificity (0.96, 95%CI: 0.89 to 1.00) of $p16^{INK4a}$ combined with an HPV DNA PCR on histological material as a test strategy. *Sources: (Prigge, 2017)*

- GRADE	Positive and negative predictive values were not reported for $p16^{INK4a}$ combined with an HPV DNA PCR on histological material as a test strategy. *Sources: (Prigge, 2017)*

Very low GRADE	There is a very low confidence in the sensitivity (range: 0.00 to 1.00) and specificity (range: 0.16 to 1.00) of the tests performed on cytologic material. *Sources: (Baldassari, 2015; Begum, 2007; Bishop, 2012; Buonocore, 2019; Chute, 2014; Hou, 2016; Jalaly, 2015; Jannapureddy, 2010; Sivars, 2017; Smith, 2014; Takes, 2016; Xu, 2016)*

Very low GRADE	There is a very low confidence in the positive predictive value (range: 0.00 to 1.00) and negative predictive value (range: 0.16 to 1.00) of the tests performed on cytologic material. *Sources: (Baldassari, 2015; Begum, 2007; Bishop, 2012; Buonocore, 2019; Chute, 2014; Hou, 2016; Jalaly, 2015; Jannapureddy, 2010; Sivars, 2017; Smith, 2014; Takes, 2016; Xu, 2016)*

Overwegingen – van bewijs naar aanbeveling

Prigge (2017) vond een hoge sensitiviteit (0,93; 95%BHI: 0,87 to 0,97; I^2: 23,4%) en een hoge specificiteit (0,96; 95%BHI: 0,89 to 1,00; I^2: 68,4%) voor het gecombineerde gebruik van $p16^{INK4a}$ met een HPV DNA PCR op histologisch materiaal van mensen met een plaveiselcarcinoom van de orofarynx. Er werden geen positief en negatief voorspellende waarden gerapporteerd. De zekerheid in deze diagnostische accuratesse werd echter als zeer laag beoordeeld, gezien het onduidelijk was of de geïncludeerde studies onwenselijke exclusies vermeden, er enige zorgen waren over de toepasbaarheid met betrekking tot te patiënten in de steekproeven en vanwege de kleine informatie grootte door het lage deelnemersaantal (imprecisie).

Er werden veel verschillende methoden en procedures gevonden voor testen op cytologisch materiaal van patiënten met een plaveiselcelcarcinoom metastase in een hals lymfklier. Zo werd er, bijvoorbeeld, getest met verschillende indextesten, verschillende referentietesten, verschillende afkapwaarden voor positieve uitslagen, verschillend materiaal voor referentietesten en/of verschillende fixatiemiddelen. Door deze heterogeniteit werd geacht

dat de data uit deze studies niet samen te voegen waren tot gepoolde schatters voor sensitiviteit en specificiteit. Hierdoor varieerden de geobserveerde sensitiviteit (range: 0,00 tot 1,00), de specificiteit (range: 0,16 tot 1,00), de positief voorspellende waarde (range: 0,00 tot 1,00) en de negatief voorspellende waarde (range: 0,00 tot 1,00), afhankelijk van de tests en procedures. De zekerheid in het gevonden bewijs was zeer laag door risico's op vertekening van uitkomsten, door enige zorgen over de toepasbaarheid en door de zeer kleine informatie grootte in elke afzonderlijke vergelijking (imprecisie).

The College of American Pathologists (CAP) ontwikkelde een richtlijn over HPV diagnostiek bij hoofd-hals carcinomen (Lewis, 2018). De CAP-richtlijn werd multidisciplinair ontwikkeld, met medische expertise, expertise op het gebied van hoofd, hals en moleculaire pathologie, en chirurgische, medische en radiatie oncologie (Lewis, 2018). Ook werd er een methodoloog aan de multidisciplinaire werkgroep toegevoegd en werd er een adviesgroep opgericht. De adviesgroep bestond uit patiëntvertegenwoordigers, pathologen, een medisch oncoloog en moleculair epidemioloog, een radiotherapeut-oncoloog en een methodoloog. Eventuele financiële belangen van de werkgroep werden in kaart gebracht. Twee (van de elf) deelnemers hadden potentiële belangen, maar specifieke acties hierop werden niet gerapporteerd. De ontwikkelmethodologie van de richtlijn werd in een supplement gerapporteerd (Lewis, 2018). Uitgangsvragen werden opgesteld en een systematisch zoekopdracht en literatuurselectie werden uitgevoerd. De in- en exclusiecriteria zijn vermeld, maar kunnen wellicht niet voor elke uitgangsvraag volledig reproduceerbaar zijn. Data werd vervolgens uit de geselecteerde studies geëxtraheerd en de studies werden op kwaliteit beoordeeld. Systematische reviews werden met de AMSTAR-tool beoordeeld en observationele studies met de Newcastle-Ottawa quality assessment scale. (Her)analyses van RCT's werden niet met een specifiek kwaliteitsinstrument beoordeeld. De richtlijnwerkgroep moest vier specifieke overwegingen maken op tot aanbevelingen te komen. De overwegingen betroffen significante bevindingen, de algehele sterkte van het bewijs, de sterkte van de te maken aanbeveling, en de balans tussen schade en voordelen. Er werd geen formeel framework gebruikt om deze beslissingen expliciet en/of transparant te maken. De CAP voorzag de werkgroep van geld voor de projectadministratie en er werden geen gelden uit de industrie gebruik. Werkgroepleden van de CAP-richtlijn werden niet gecompenseerd voor hun betrokkenheid en investeerden kosteloos hun tijd.

De CAP-richtlijn rapporteert een algoritme voor de work-up van patiënt monster (Lewis, 2018). Het algoritme start met een monster door biopsie of resectie van een gediagnosticeerd plaveiselcarcinoom en vertakt afhankelijk van de tumorlocatie (i.e. multi-site met een betrokken oropharynx, cervicale lymfklier, non-orofaryngeale primaire tumor, en orofaryngeale tumor). In het algoritme is de eerste test uit de work-up p16 immunohistochemie wanneer een HPV test geïndiceerd is. Hierin wordt ≥ 70% kleuring van de nuclei en cytoplasma als een positief resultaat gezien. Alleen wanneer het carcinoom keratinizerend is, de metastase zich niet in level II of level III van de hals bevindt, en/of wanneer de betrokken lymfklier niet bepaald kan worden stellen de CAP-richtlijn auteurs dat een HR HPV-specifieke test noodzakelijk is om een positieve p16 immunohistochemische test te bevestigen (Lewis, 2018). De diagnostische accuratesse van het voorgestelde algoritme werd niet onderzocht. Figuur 1 in de CAP-richtlijn laat het gehele voorgestelde algoritme zien van de diagnostische work-up (Lewis, 2018). De richtlijn werd als een open access artikel gepubliceerd (zie de URLs in de bijlagen van deze module).

De CAP-richtlijn vermeldde, als aanbeveling, dat op materiaal van patiënten met een cervicale plaveiselcelcarcinoom metastase van een onbekend primair carcinoom routinematig HR HPV testen zou moeten worden gedaan (Lewis, 2018). Deze vermelding als

aanbeveling betekent dat er enkele limitaties aan de kwaliteit van het bewijs, balans tussen schade en voordelen, waarden of kosten zitten. Verder werd er een expert consensus uitspraak gedaan over HR HPV tests op materiaal afgenomen via een fijne naald aspiraat bij patiënten met een plaveiselcarcinoom metastase van een onbekende primaire tumor. De consensus onder de experts in de CAP-richtlijnwerkgroep was dat HR HPV tests bij alle patiënten met plaveiselcelcarcinoom metastasen van een onbekende primaire tumor zouden moeten worden uitgevoerd (Lewis, 2018). Een expert consensus uitspraak in de CAP-richtlijn betekent dat de werkgroep van de CAP het noodzakelijk achtte om over het onderwerp een uitspraak te doen, maar dat er ernstige limitaties zijn aan de kwaliteit van het bewijs, balans tussen schade en voordelen, waarden of kosten. De CAP-richtlijn vermeldde verder dat er geen specifieke aanbevelingen gegeven konden worden over de test methodologie en dat testen op weefsel (wanneer beschikbaar) uitgevoerd zouden moeten worden indien de HR HPV test op het cytologische materiaal negatief was (Lewis, 2018). Verder merkt de CAP-richtlijnwerkgroep op dat pathologen de afkapwaarden (voor positieve/negatieve uitslagen) zouden moeten valideren wanneer men p16 immunohistochemie op cytologisch materiaal gebruikt (Lewis, 2018).

Uit het literatuuronderzoek zijn geen eenduidige aanbevelingen te destilleren met betrekking tot de “beste” test of combinatie van testen om HR-HPV aan te tonen. Ook het literatuuronderzoek dat verricht is voor het schrijven van de CAP richtlijn heeft dit niet kunnen aantonen. Derhalve sluit de werkgroep zich aan bij de aanbevelingen uit de CAP richtlijn die wel beschrijven in welke situaties er wel en niet getest moet worden voor HPV door de patholoog, maar niet expliciet voorschrijven welke test daarvoor gebruikt moet worden (Lewis, 2018). De makkelijke beschikbaar, relatieve eenvoud en brede beschikbaarheid van p16 immunohistochemie geldt daarbij wel als minimum basis wat elk laboratorium moet kunnen uitvoeren.

Het doel van de PA diagnose is om met het beschikbare materiaal de beste diagnose voor de patiënt te stellen. Voor patiënten is het van belang dat de tumor juist wordt geclassificeerd voor de best mogelijke behandeling. Hierbij is het bepalen van de HPV-status van belang, omdat deze tumor een aparte classificatie heeft gekregen in de TNM 8^{e} editie. Eventueel kan de hoog-risico HPV nog nader getypeerd worden.

P16 immunohistochemie is een relatief simpele en snel uit te voeren test die elk PA laboratorium in Nederland standaard in zijn pakket heeft. Voor de overige (moleculaire) technieken zijn er verschillen welke test ter beschikking is. Alle PA laboratoria in Nederland zijn ISO15189 geaccrediteerd en voeren dus geregeld kwaliteitscontroles en interlaboratorium vergelijkingen uit voor al hun beschikbare testen, zodat onafhankelijk van het platform de kwaliteit van de uitslag geborgd is.

Aanbevelingen

Aanbeveling-1

Rationale van de aanbeveling: weging van argumenten voor en tegen de interventies

De aanbevelingen zijn met enige aanpassingen voor de Nederlandse situatie aangenomen vanuit de richtlijn van de College of American Pathologists (Lewis, 2018). Kennis van het rookgedrag van de patiënt mag het uitvoeren van de HPV-test niet beïnvloeden.

Onafhankelijkheid

Voer een HPV-test onafhankelijk van kennis over het anamnestische rookgedrag van de patiënt uit.

Aanbeveling-2

Rationale van de aanbeveling: weging van argumenten voor en tegen de interventie

De aanbevelingen zijn met enige aanpassingen voor de Nederlandse situatie aangenomen vanuit de richtlijn van de College of American Pathologists (Lewis, 2018). Het is belangrijk om, daar waar mogelijk, op histologisch materiaal te testen. Voor het gebruik van p16 immunohistochemie als eerste test op histologisch materiaal werd gekozen omdat deze test makkelijk beschikbaar en relatieve eenvoudig uit te voeren is. Indien er geen histologisch materiaal beschikbaar is of beschikbaar komt, is het testen op cytologisch materiaal van een lymfklierpunctaat een alternatief. Het is onduidelijk vanuit de literatuur met welke test er op cytologisch materiaal gebruikt zou moeten worden.

Bij (klinische verdenking op) nieuwe orofarynxcarcinomen

- Voer een HR-HPV test uit op alle nieuw gediagnosticeerde plaveiselcelcarcinomen van de orofarynx , onafhankelijk van het histologische subtype.
- Voer de HR-HPV test uit op de primaire tumor of op een metastase indien deze metastase klinisch afkomstig is van het orofarynxcarcinoom.
- Voer p16 immunohistochemie uit op histologisch materiaal van een orofarynxcarcinoom als HR-HPV test. Overweeg een additionele specifieke test als bevestiging.
- Voer HR-HPV testen uit op cytologisch materiaal van een lymfklierpunctaat indien er geen histologisch materiaal aanwezig is en histologisch materiaal niet te verkrijgen is bij patiënten met een orofarynxcarcinoom dat nog niet eerder getest is of bij patiënten met een klinische verdenking op een orofarynxcarcinoom.

Aanbeveling-3

Rationale van de aanbeveling: weging van argumenten voor en tegen de interventie

De aanbevelingen zijn met enige aanpassingen voor de Nederlandse situatie aangenomen vanuit de richtlijn van de College of American Pathologists (Lewis, 2018). Indien er histologisch materiaal beschikbaar is uit een Level II of III lymfekliermetastase werd er voor p16 als eerste test gekozen omdat deze test makkelijk beschikbaar en relatieve eenvoudig uit te voeren is. Omdat er in deze situatie niet altijd histologisch materiaal beschikbaar is, worden testen op cytologisch materiaal uit een lymfklierpunctaat als alternatief gezien. Het is onduidelijk vanuit de literatuur met welke test er op cytologisch materiaal gebruikt zou moeten worden.

Bij metastasen van onbekende primaire tumor

- Voer routinematig een HR-HPV test uit op materiaal van patiënten met een plaveiselcelcarcinoommetastase van onbekende primaire tumor bij metastasen in Level II of III van de hals.
- Voer p16 immunohistochemie uit op histologisch materiaal uit een Level II of III lymfekliermetastase met onbekende primaire tumor.
- Voer p16 immunohistochemie uit op histologisch materiaal van lymfekliermetastase buiten LII of LIII met onbekende primaire tumor in geval van niet-keratiniserende morfologie. Overweeg een additionele specifieke test als bevestiging.
- Voer HR-HPV testen uit op cytologisch materiaal van een lymfklierpunctaat indien er geen histologisch materiaal aanwezig is en dit materiaal niet te verkrijgen is bij patiënten met een onbekende primaire tumor.

NB: er wordt geen aanbeveling gegeven over de te gebruiken test op cytologisch materiaal.

Aanbeveling-4

Rationale van de aanbeveling: weging van argumenten voor en tegen de interventie

De aanbevelingen zijn met enige aanpassingen voor de Nederlandse situatie aangenomen vanuit de richtlijn van de College of American Pathologists (Lewis, 2018). Routinematige HR-HPV testen voor niet-plaveiselcelcarcinomen, andere primaire hoofd-halscarcinomen, en residu van of recidiverende eerste tumoren (waarvan de HPV status al initieel bepaald werd) worden niet aangeraden. Voor plaveiselcelcarcinomen van het hoofd-halsgebied werd routinematige laag risico HPV (LR-HPV) testen niet aanbevolen.

Niet routinematig onderzoek

- Voer niet routinematig HR-HPV testen uit voor niet-plaveiselcelcarcinomen.
- Voer niet routinematig HR-HPV testen uit op andere primaire hoofd-hals carcinomen dan orofarynx.
- Overweeg geen HR-HPV testen uit te voeren bij patiënten met een residu of recidiverende tumor waarvan de HPV status initieel al was vastgesteld. Overweeg bij twijfel of het een recidiverende eerste tumor is om wél een HR-HPV test uit te voeren.
- Voer niet routinematig laag risico HPV testen uit op plaveiselcelcarcinomen van het hoofd-halsgebied.

Aanbeveling-5

Rationale van de aanbeveling: weging van argumenten voor en tegen de interventie

De aanbevelingen zijn met enige aanpassingen voor de Nederlandse situatie aangenomen vanuit de richtlijn van de College of American Pathologists (Lewis, 2018).

Rapportage

- Rapporteer p16-positiviteit in het histologisch materiaal bij ten minste matige tot sterke aankleuring van 70% van de cellen als surrogaat voor HR-HPV.
- Rapporteer p16 immunohistochemie-positieve of HR-HPV-positieve primaire orofarynxcarcinomen als p16-positief of HPV-positief.
- Gradeer de HPV/p16-positieve orofarynxcarcinomen niet.

Literatuur

Baldassarri R, Aronberg R, Levi AW, Yarbrough WG, Kowalski D, Chhieng D. Detection and genotype of high-risk human papillomavirus in fine-needle aspirates of patients with metastatic squamous cell carcinoma is helpful in determining tumor origin. Am J Clin Pathol. 2015 May;143(5):694-700. doi: 10.1309/AJCPCZA4PSZCFHQ4. PMID: 25873503.

Begum S, Gillison ML, Nicol TL, Westra WH. Detection of human papillomavirus-16 in fine-needle aspirates to determine tumor origin in patients with metastatic squamous cell carcinoma of the head and neck. Clin Cancer Res. 2007 Feb 15;13(4):1186-91. doi: 10.1158/1078-0432.CCR-06-1690. PMID: 17317828.

Bishop JA, Maleki Z, Valsamakis A, Ogawa T, Chang X, Pai SI, Westra WH. Application of the hybrid capture 2 assay to squamous cell carcinomas of the head and neck: a convenient liquid-phase approach for the reliable determination of human papillomavirus status. Cancer Cytopathol. 2012 Feb 25;120(1):18-25. doi: 10.1002/cncy.20175. Epub 2011 Jul 12. PMID: 21751428; PMCID: PMC3197962.

Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, Fervers B, Graham ID, Grimshaw J, Hanna SE, Littlejohns P, Makarski J, Zitzelsberger L; AGREE Next Steps Consortium. AGREE II: advancing guideline development, reporting and evaluation in health care. CMAJ. 2010 Dec 14;182(18):E839-42. doi: 10.1503/cmaj.090449. Epub 2010 Jul 5. PMID: 20603348; PMCID: PMC3001530.

Buonocore DJ, Fowle E, Lin O, Xu B, Katabi N, Cohen JM. Cytologic evaluation of p16 staining in head and neck squamous cell carcinoma in CytoLyt versus formalin-fixed material. Cancer Cytopathol. 2019 Dec;127(12):750-756. doi: 10.1002/cncy.22191. Epub 2019 Oct 10. PMID: 31600033; PMCID: PMC6906234.

Chute DJ, Aramouni GT, Brainard JA, Hoschar AP, Kroeger A, Yen-Lieberman B. Hybrid Capture 2 human papilloma virus testing for head and neck cytology specimens. *J Am Soc Cytopathol*. 2014;3(4):173-182. doi:10.1016/j.jasc.2014.02.004.

Hou Y, Chaudhary S, Shen R, Li Z. Fine-needle aspiration of cervical lymph nodes yields adequate materials for accurate HPV testing in metastatic head and neck squamous cell carcinomas. Diagn Cytopathol. 2016 Oct;44(10):792-8. doi: 10.1002/dc.23548. Epub 2016 Jul 28. PMID: 27465660.

Jalaly JB, Lewis JS Jr, Collins BT, Wu X, Ma XJ, Luo Y, Bernadt CT. Correlation of p16 immunohistochemistry in FNA biopsies with corresponding tissue specimens in HPV-related squamous cell carcinomas of the oropharynx. Cancer Cytopathol. 2015 Dec;123(12):723-31. doi: 10.1002/cncy.21600. Epub 2015 Aug 4. PMID: 26242494.

Jannapureddy S, Cohen C, Lau S, Beitler JJ, Siddiqui MT. Assessing for primary oropharyngeal or nasopharyngeal squamous cell carcinoma from fine needle aspiration of cervical lymph node metastases. Diagn Cytopathol. 2010 Nov;38(11):795-800. doi: 10.1002/dc.21293. PMID: 20014308.

Lewis JS Jr, Beadle B, Bishop JA, Chernock RD, Colasacco C, Lacchetti C, Moncur JT, Rocco JW, Schwartz MR, Seethala RR, Thomas NE, Westra WH, Faquin WC. Human Papillomavirus Testing in Head and Neck Carcinomas: Guideline From the College of American Pathologists. Arch Pathol Lab Med. 2018 May;142(5):559-597. doi: 10.5858/arpa.2017-0286-CP. Epub 2017 Dec 18. PMID: 29251996.

Prigge ES, Arbyn M, von Knebel Doeberitz M, Reuschenbach M. Diagnostic accuracy of $p16^{INK4a}$ immunohistochemistry in oropharyngeal squamous cell carcinomas: A systematic review and meta-analysis. Int J Cancer. 2017 Mar 1;140(5):1186-1198. doi: 10.1002/ijc.30516. Epub 2016 Dec 2. PMID: 27859245.

Sivars L, Landin D, Haeggblom L, Tertipis N, Grün N, Bersani C, Marklund L, Ghaderi M, Näsman A, Ramqvist T, Nordfors C, Munck-Wikland E, Tani E, Dalianis T. Human papillomavirus DNA detection in fine-needle aspirates as indicator of human papillomavirus-positive oropharyngeal squamous cell carcinoma: A prospective study. Head Neck. 2017 Mar;39(3):419-426. doi: 10.1002/hed.24641. Epub 2016 Nov 29. PMID: 27898186.

Smith DF, Maleki Z, Coughlan D, Gooi Z, Akpeng B, Ogawa T, Bishop JA, Frick KD, Agrawal N, Gourin CG, Ha PK, Koch WM, Richmon JD, Westra WH, Pai SI. Human papillomavirus status of head and neck cancer as determined in cytologic specimens using the hybrid-capture 2 assay. Oral Oncol. 2014 Jun;50(6):600-4. doi: 10.1016/j.oraloncology.2014.02.011. Epub 2014 Mar 12. PMID: 24630260; PMCID: PMC4318229.

Takes RP, Kaanders JH, van Herpen CM, Merkx MA, Slootweg PJ, Melchers WJ. Human papillomavirus detection in fine needle aspiration cytology of lymph node metastasis of head and neck squamous cell cancer. J Clin Virol. 2016 Dec;85:22-26. doi: 10.1016/j.jcv.2016.10.008. Epub 2016 Oct 27. PMID: 27816020.

Xu B, Ghossein R, Lane J, Lin O, Katabi N. The utility of p16 immunostaining in fine needle aspiration in p16-positive head and neck squamous cell carcinoma. Hum Pathol. 2016 Aug;54:193-200. doi: 10.1016/j.humpath.2016.04.002. Epub 2016 Apr 19. PMID: 27105759.

Geldigheid en Onderhoud

Module[1]	Regiehouder(s)[2]	Jaar van autorisatie	Eerstvolgende beoordeling actualiteit richtlijn[3]	Frequentie van beoordeling op actualiteit[4]	Wie houdt er toezicht op actualiteit[5]	Relevante factoren voor wijzigingen in aanbeveling[6]

[1] Naam van de module
[2] Regiehouder van de module (deze kan verschillen per module en kan ook verdeeld zijn over meerdere regiehouders)
[3] Maximaal na vijf jaar
[4] (half)Jaarlijks, eens in twee jaar, eens in vijf jaar
[5] regievoerende vereniging, gedeelde regievoerende verenigingen, of (multidisciplinaire) werkgroep die in stand blijft
[6] Lopend onderzoek, wijzigingen in vergoeding/organisatie, beschikbaarheid nieuwe middelen

Bijlagen bij module 2.1

Kennislacunes
What is the diagnostic accuracy of diagnostic test algorithms to determine the HPV-status on histological material in patients with an oropharyngeal carcinoma?

What is the diagnostic accuracy of cytologic tests in patients with a carcinoma of unknown primary?

Indicatoren
Geen.

Implementatieplan

Aanbeveling	Tijdspad voor implementatie: < 1 jaar, 1 tot 3 jaar of 3 tot 5 jaar	Verwacht effect op kosten	Randvoorwaarden voor implementatie (binnen aangegeven tijdspad)	Mogelijke barrières voor implementatie[1]	Te ondernemen acties voor implementatie[2]	Verantwoordelijken voor acties[3]	Overige opmerkingen
Onafhankelijkheid: Voer de HPV-testen onafhankelijk van kennis over het anamnestische rookgedrag van de patiënt uit.	< 1 jaar	Geen	Geen	Geen	Bekendheid van de richtlijn(module) vergroten in het beroepenveld (bijvoorbeeld door middel van nieuwsberichten of presentaties)	NVVP	Geen
Bij (klinische verdenking op) nieuwe orofarynxcarcinomen: • Voer een HR-HPV test uit op alle nieuw gediagnosticeerde plaveiselcelcarcinomen van de orofarynx , onafhankelijk van het histologische subtype. • Voer de HR-HPV test uit op de primaire tumor of op een metastase indien deze metastase klinisch afkomstig is van	< 1 jaar	Geen	Geen	Geen	Bekendheid van de richtlijn(module) vergroten in het beroepenveld (bijvoorbeeld door middel van nieuwsberichten of presentaties)	NVVP	Geen

het orofarynxcarcinoom. • Voer p16 immunohistochemie uit op histologisch materiaal van een orofarynxcarcinoom als HR-HPV test. Overweeg een additionele specifieke test als bevestiging. • Voer HR-HPV testen uit op cytologisch materiaal van een lymfklierpunctaat indien er geen histologisch materiaal aanwezig is en histologisch materiaal niet te verkrijgen is bij patiënten met een orofarynxcarcinoom dat nog niet eerder getest is of bij patiënten met een klinische verdenking op een orofarynxcarcinoom.							
Bij metastasen van onbekende primaire tumor: • Voer routinematig een HR-HPV test uit op materiaal van patiënten met een plaveiselcelcarcinoommetastase van onbekende primaire tumor bij metastasen in Level II of III van de hals. • Voer p16 immunohistochemie uit op histologisch materiaal uit een Level II of III lymfekliermetastase met onbekende primaire tumor. • Voer p16 immunohistochemie uit op histologisch	< 1 jaar	Geen	Geen	Geen	Bekendheid van de richtlijn(module) vergroten in het beroepenveld (bijvoorbeeld door middel van nieuwsberichten of presentaties)	NVVP	Geen

materiaal van lymfekliermetastase buiten LII of LIII met onbekende primaire tumor in geval van niet-keratiniserende morfologie.. Overweeg een additionele specifieke test als bevestiging. • Voer HR-HPV testen uit op cytologisch materiaal van een lymfklierpunctaat indien er geen histologisch materiaal aanwezig is en dit materiaal niet te verkrijgen is bij patiënten met een onbekende primaire tumor. NB: er wordt geen aanbeveling gegeven over de te gebruiken test op cytologisch materiaal.							
Niet routinematig onderzoek: • Voer niet routinematig HR-HPV testen uit voor niet-plaveiselcelcarcinomen. • Voer niet routinematig HR-HPV testen uit op andere primaire hoofd-hals carcinomen dan orofarynx. • Overweeg geen HR-HPV testen uit te voeren bij patiënten met een residu of recidiverende tumor waarvan de HPV status initieel al was vastgesteld. Overweeg bij twijfel of het een recidiverende eerste tumor is om wél een HR-HPV test uit te voeren.	< 1 jaar	Geen	Geen	Geen	Bekendheid van de richtlijn(module) vergroten in het beroepenveld (bijvoorbeeld door middel van nieuwsberichten of presentaties)	NVVP	Geen

• Voer niet routinematig laag risico HPV testen uit op plaveiselcelcarcinomen van het hoofd-halsgebied.							
Rapportage: • Rapporteer p16-positiviteit in het histologisch materiaal bij ten minste matige tot sterke aankleuring van 70% van de cellen als surrogaat voor HR-HPV. • Rapporteer p16 immunohistochemie-positieve of HR-HPV-positieve primaire orofarynxcarcinomen als p16-positief of HPV-positief. • Gradeer de HPV/p16-positieve orofarynxcarcinomen niet.	< 1 jaar	Geen	Geen	Geen	Opname in de PALGA protocol module, waarvoor overleg met Stichting PALGA nodig is	NVVP	Geen

[1] Barrières kunnen zich bevinden op het niveau van de professional, op het niveau van de organisatie (het ziekenhuis) of op het niveau van het systeem (buiten het ziekenhuis). Denk bijvoorbeeld aan onenigheid in het land met betrekking tot de aanbeveling, onvoldoende motivatie of kennis bij de specialist, onvoldoende faciliteiten of personeel, nodige concentratie van zorg, kosten, slechte samenwerking tussen disciplines, nodige taakherschikking, et cetera.
[2] Denk aan acties die noodzakelijk zijn voor implementatie, maar ook acties die mogelijk zijn om de implementatie te bevorderen. Denk bijvoorbeeld aan controleren aanbeveling tijdens kwaliteitsvisitatie, publicatie van de richtlijn, ontwikkelen van implementatietools, informeren van ziekenhuisbestuurders, regelen van goede vergoeding voor een bepaald type behandeling, maken van samenwerkingsafspraken.
[3] Wie de verantwoordelijkheden draagt voor implementatie van de aanbevelingen, zal tevens afhankelijk zijn van het niveau waarop zich barrières bevinden. Barrières op het niveau van de professional zullen vaak opgelost moeten worden door de beroepsvereniging. Barrières op het niveau van de organisatie zullen vaak onder verantwoordelijkheid van de ziekenhuisbestuurders vallen. Bij het oplossen van barrières op het niveau van het systeem zijn ook andere partijen, zoals de NZA en zorgverzekeraars, van belang.

Evidence tables

Evidence table for algorithms (PICO 1)

Study reference	Study characteristics	Patient characteristics	Index test (test of interest)	Reference test	Follow-up	Outcome measures and effect size	Comments
Prigge 2017 (individual study characteristics deduced from Prigge 2017)	SR and meta-analysis *Literature searches up to the 8th of January 2016.* Included articles for p16INK4a + HPV DNA PCR **A**: Smeets 2007 **B**: Schache 2011 **C**: Schlecht 2011 **D**: Rotnaglova 2011 **E**: Hoffmann 2012 **F**: Rietbergen 2013 **G**: Bussu 2013 **H**: Bussu 2014 **I**: Lukesova 2014 **J:** Masterson 2015 **K:** Vojtechova 2016 Study design: prospective and retrospective designs were included (unclear whether case-control designs could be included. Specific designs not reported in the SR) Setting and Country: **A**: Netherlands **B**: UK	Inclusion criteria SR: men and women diagnosed with oropharyngeal squamous cell carcinoma, p16INK4a IHC as index test, A reference test that would detect E6 and/or E7 HPV mRNA, sensitivity and specificity as outcomes, prospective or retrospective studydesign. Exclusion criteria SR: No author response when inquiry was made about data, samplesize <10, no primary data. *11 studies included for p16INK4a + HPV DNA PCR (24 studies in total)* Important patient characteristics: **Patient characteristics are not reported by the**	Describe index and comparator tests* and cut-off point(s): All studies described here used p16INK4a + HPV DNA PCR as an index test. Cut-off points (p16INK4a + HPV DNA PCR): Positive: p16INK4a and HPV DNA PCR both positive Negative: Either one or both of the tests (i.e., p16INK4a and/or HPV DNA PCR) was negative. Cut-off points for p16INK4a: **A**: Staining intensity above background of negative control **B**: ≥70% strong and diffuse nuclear and cytoplasmic staining	Describe reference test and cut-off point(s): The exact reference test per study is unclear from the systematic review. Nonetheless, it was a test for E6 and/or E7 mRNA. Detection HPV-DNA (transcript types / HPV types): **A**: E6*I / HPV 16 **B**: E6 / HPV 16 **C**: E6, E7, E6*I / HPV 16 **D**: E6*I / HPV 16 **E**: E6*I / HPV 16, 18, 26, 31, 33, 35, 39, 45, 51, 52, 53, 56,	Endpoint of follow-up: **N/A**	Outcome measures and effect size (include 95%CI and p-value if available): Sensitivity P16INK4a + HPV DNA PCS versus a reference test detecting E6 and/or E7 mRNA: sensitivity (95%CI) **A**: 1.00 (0.57-1.00) **B**: 0.96 (0.82-0.99) **C**: 0.91 (0.76-0.97) **D**: 0.89 (0.57-0.98) **E**: 0.82 (0.52-0.95) **F**: 0.96 (0.80-0.99) **G**: 0.70 (0.40-0.89)	Study quality (ROB): QUADAS-2 Place of the index test in the clinical pathway: Unclear Choice of cut-off point: Various cut-off points were observed for p16INK4a. Various HPV genes were targeted. Facultative: Brief description of author's conclusion Personal remarks on study quality, conclusions, and other issues (potentially) relevant to the research question Sensitivity analyses (excluding small studies; excluding low quality studies; excluding case-control type of studies; relevant subgroup-analyses); mention only analyses which are of

	C: USA **D**: Czech Republic **E**: Germany **F**: Netherlands **G**: Italy **H**: Italy **I**: Czech Republic **J:** UK **K:** Czech Republic <u>Source of funding and conflicts of interest:</u> One author is a co-inventor of various patents regarding the diagnostic use op p16INK4a antibodies / was co-founder, shareholder and member of a company that developed and marketed p16INK4a related reagents (later aquired by Roche) / holds a patent regarding therapeutic use of p16INK4a / received honoraria as a scientific advisor and received research funds frum Oryx GmbH & Co. The salary of another author was partially funded by the research funds received from Oryx GmbH & Co. Both authors are inventors of a patent related to the therapeutic use of p16INK4a.	**systematic review authors.** <u>Sample size, n:</u> **A**: 15 **B**: 82 **C**: 19 **D**: 45 **E**: 20 **F**: 86 **G**: 21 **H**: 22 **I**: 52 **J:** 24 **K:** 123	**C**: Mean intensity of ≥2 and ≥75% staining in either the nuclei or cytoplasm **D**: >50% nuclear and/or cytoplasmic staining **E**: strong nuclear and cytoplasmic staining in focal or diffuse distribution **F**: >70% moderate to strong diffuse nuclear and cytoplasmic staining **G**: ≥70% strong and diffuse nuclear and cytoplasmic staining **H**: ≥70% strong and diffuse nuclear and cytoplasmic staining **I**: >50% nuclear and/or cytoplasmic staining **J:** >70% staining **K:** >50% nuclear and/or cytoplasmic staining Material for p16INK4a: **A**: FFPE (whole tissue section) **B**: FFPE (tissue microarray) **C**: FFPE (whole tissue section) **D**: FFPE (whole tissue section	58, 59, 66, 67, 68b, 70, 73, 82 **F**: E6*I, E6 / HPV 16, 33 **G**: E6, E7 / HPV 16, 18, 31, 33, 45 **H**: E6, E7 / HPV 16, 18, 31, 33, 35 **I**: E6*I / HPV 16 **J:** E6, E7 / HPV 16 **K:** E6*I / HPV 16 Material for reference test: **A**: FFPE (whole tissue section) **B**: Fresh frozen **C**: Fresh frozen **D**: FFPE (whole tissue section **E**: Fresh frozen **F**: Fresh frozen **G**: Fresh frozen **H**: Fresh frozen **I**: FFPE (whole tissue section) **J:** Fresh frozen **K:** FFPE (whole tissue section) Prevalence (%) (based on refence test at		**H**: 0.82 (0.52-0.95) **I**: 0.81 (0.63-0.92) **J:** 1.00 (0.70-1.00) **K:** 0.96 (0.88-0.99) Pooled characteristic (bivariate analysis) per index test and cut-off point: **Index test (various cut-offs)** 0.93 (95% CI 0.87 to 0.97) Heterogeneity (reasons): I2 = 23.39% (p=0.22) <u>Specificity P16INK4a + HPV DNA PCS versus a reference test detecting E6 and/or E7 mRNA:</u> Specificity (95%CI) **A**: 1.00 (0.72-1.00) **B**: 0.94 (0.74-0.99)	potential importance to the research question. Heterogeneity: clinical and statistical heterogeneity; clinical: enough similarities in patient characteristics, diagnostic tests (strategy) to allow pooling? For pooled data: assessment of statistical heterogeneity and, more importantly, assessment of the reasons for heterogeneity (if present)? Note: sensitivity and specificity depend on the situation in which the test is being used and the thresholds that have been set, and sensitivity and specificity are correlated; therefore, the use of heterogeneity statistics (p-values; I2) is problematic, and rather than testing whether heterogeneity is present, the reasons for heterogeneity should be examined. **A**: **B**: **C**: **D**: **E**: **F**: **G**: **H**:

			E: FFPE (whole tissue section) **F**: FFPE (whole tissue section) **G**: FFPE (whole tissue section) **H**: FFPE (whole tissue section) **I**: FFPE (whole tissue section) **J:** FFPE (whole tissue section) **K:** FFPE (whole tissue section) Antibody clone p16INK4a: **A**: E6H4 **B**: E6H4 **C**: G175-405 **D**: G175-405 **E**: E6H4 **F**: E6H4 **G**: E6H4 **H**: E6H4 **I**: G175-405 **J:** G175-405 **K:** G175-405 ..... HPV-DNA PCR method: **A**: GP5+/6+ reverse line blot genotyping **B**: HPV16 primers **C**: MY09/MY11/HMB01	specified cut-off point) **Not reported in the SR** For how many participants were no complete outcome data available? **Not reported in the SR** Reasons for incomplete outcome data described. **Not reported in the SR**		**C**: 0.94 (0.83-0.98) **D**: 1.00 (0.72-1.00) **E**: 1.00 (0.70-1.00) **F**: 0.98 (0.91-1.00) **G**: 1.00 (0.74-1.00) **H**: 1.00 (0.74-1.00) **I**: 1.00 (0.87-1.00) **J:** 0.53 (0.30-0.75) **K:** 0.81 (0.68-0.89) Pooled characteristic (bivariate analysis) per index test and cut-off point: **Index test (various cut-offs)** 0.96 (95% CI 0.89 to 1.00) Heterogeneity (reasons): I2= 68.4% (p=0.00)	**I**: **J:** **K:** .....

			dot blot hybridization genotyping **D**: GP5+/6+ reverse line blot genotyping **E**: BSGP5+/6+ bead-based genotyping **F**: GP5+/6+ and EIA readout with bead-based genotyping **G**: Hybrid capture 2 **H**: Hybrid capture 2 **I**: PF5+/6+ reverse line blot genotyping **J:** HPV16 E6/E7 primers **K:** GP5+/6+ reverse line blot genotyping				

***comparator test equals the C of the PICO; two or more index/ comparator tests may be compared; note that a comparator test is not the same as a reference test (golden standard)**

Table of quality assessment for systematic reviews of diagnostic studies

Based on AMSTAR checklist (Shea, 2007; BMC Methodol 7: 10; doi:10.1186/1471-2288-7-10) and PRISMA checklist (Moher, 2009; PLoS Med 6: e1000097; doi:10.1371/journal. pmed1000097)

Research question:

Study **First author, year**	**Appropriate and clearly focused question?[1]** **Yes/no/unclear**	**Comprehensive and systematic literature search?[2]** **Yes/no/unclear**	**Description of included and excluded studies?[3]** **Yes/no/unclear**	**Description of relevant characteristics of included studies?[4]** **Yes/no/unclear**	**Assessment of scientific quality of included studies?[5]** **Yes/no/unclear**	**Enough similarities between studies to make combining them reasonable?[6]** **Yes/no/unclear**	**Potential risk of publication bias taken into account?[7]** **Yes/no/unclear**	**Potential conflicts of interest reported?[8]** **Yes/no/unclear**
Prigge	Yes, PICO was also provided in the supplement	Yes, search string was described. MEDLINE was searched.	No, no references were made to the (full-text) excluded articles.	No, sample characteristics were not described. For testprocedures, most relevant characteristics are described (see supplement as well). However, cut-off point in the reference test and methods of the reference test should have been described as well.	Yes, QUADAS-2 tool was used.	Unclear, differences in cut-offs from the index/comparator tests were analysed in sub-analyses. Unclear what the impact of the variability in cut-offs is compared to eachother.	No, publication bias was not discussed.	No, the authors disclose their conflicts of interest. CoI of included studies were not reported.

1. **Research question (PICO) and inclusion criteria should be appropriate (in relation to the research question to be answered in the clinical guideline) and predefined.**
2. **Search period and strategy should be described; at least Medline searched.**
3. **Potentially relevant studies that are excluded at final selection (after reading the full text) should be referenced with reasons.**
4. **Characteristics of individual studies relevant to the research question (PICO) should be reported.**
5. **Quality of individual studies should be assessed using a quality scoring tool or checklist (preferably QUADAS-2; COSMIN checklist for measuring instruments) and taken into account in the evidence synthesis.**
6. **Clinical and statistical heterogeneity should be assessed; clinical: enough similarities in patient characteristics, diagnostic tests (strategy) to allow pooling? For pooled data: at least 5 studies available for pooling; assessment of statistical heterogeneity and, more importantly (see Note), assessment of the reasons for heterogeneity (if present)? Note: sensitivity and specificity depend on the situation in which the test is being used and the thresholds that have been set, and sensitivity and specificity are correlated; therefore, the use of heterogeneity statistics (p-values; I^2) is problematic, and rather than testing whether heterogeneity is present, heterogeneity should be assessed by eye-balling (degree of overlap of confidence intervals in Forest plot), and the reasons for heterogeneity should be examined.**

7. **There is no clear evidence for publication bias in diagnostic studies, and an ongoing discussion on which statistical method should be used. Tests to identify publication bias are likely to give false-positive results, among available tests, Deeks' test is most valid. Irrespective of the use of statistical methods, you may score "Yes" if the authors discuss the potential risk of publication bias.**
8. **Sources of support (including commercial co-authorship) should be reported in both the systematic review and the included studies. Note: To get a "yes," source of funding or support must be indicated for the systematic review AND for each of the included studies.**

Evidence table for cytologic testing (PICO2)

Study reference	Study characteristics	Patient characteristics	Index test (test of interest)	Reference test	Follow-up	Outcome measures and effect size	Comments
Baldassari 2015	Type of study: Prospective Setting and country: University medical center, USA Funding and conflicts of interest: funding and COI are not reported in the manuscript	Inclusion criteria: Unclear Exclusion criteria: Unclear N=37 with 42 FNAs Prevalence: Unclear, FNAs were tested, not patients. (20/41 FNAs were positive, 1 invalid result by the reference) Mean age ± SD: 60.4 (11.8) Sex: 31M / 6F Other important characteristics: History of HNSCC, n: 17 FNA location:	Describe index test: cobas 4800 in cytologic material. FNA was performed with a 25-gauge needle (no imaging guidance). Direct air-dried and alcohol fixed smears were prepared and stained (Diff-Quik stain and Papanicolaou stain). One or two drops of the cell suspension was added to prepare a ThinPrep slide after centrifugation. The assay was performed according to the manufacturer's protocol. Cut-off point(s): Unclear, interpretation was carried out using proprietary software (provided with cobas z 480 analyzer). HPV 16, 18, pooled (31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68) DNA detection by	Describe reference test: P16 IHC and HPV ISH on histologic samples (FFPE material). P16: 5-micrometer sections were deparaffinized. Sections were incubated with a prediluted monoclinal antibody (mouse, E6H4) for 32 minutes. HPV ISH: sections were incubated with INFORM HPV III family 16 probe for 4 minutes. Probes detected HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66. Cut-off point(s): P16: unclear HPV ISH: unclear	Time between the index test en reference test: Unclear For how many participants were no complete outcome data available? Seven specimen were either atypical or suspicious in cytology. 1 cytologic atypical was negative (reference) and 2 were positive (reference). 1 cytological suspicious was negative (reference) and 3 were positive (reference) One specimen was	Outcome measures and effect size (include 95%CI and p-value if available): Cobas 4800 versus p16+ISH, n=34: TP: 15 FP: 16 FN: 0 TN: 3 Excluded: 3 atypical cytological results, 4 cytological suspicious results, 1 invalid reference result Sensitivity: 1.00 (0.78-1.00) Specificity: 0.16 (0.03-0.40)	

		Neck mass: 36 Mediastinal lymph nodes: 5 Left parapharyngeal mass: 1	amplification. Cycle thresholds <40.5 for HPV 16, <40 HPV 18 and pooled genotypes.		invalid for reference testing (positive for cytology). Reasons for incomplete outcome data described? Not reported.	*sens/spec calculated from presented data	
Begum 2007	Type of study: Database Setting and country: Hospital, USA Funding and conflicts of interest: funding and COI are not reported in the manuscript	Inclusion criteria: when initial processing of the FNA included preparation of a cell block, if the preparation of the cell block was spun in a cellular pellet Exclusion criteria: Unclear (hpv was detected in 13/77 FNAs) N= 19 Prevalence: Unclear, not all n=19 had reference testing. Mean age ± SD: Sex: % M / % F Other important characteristics:	Describe index test: P16: 5-micron sections were deparaffinized. Sections were incubated with mouse monoclonal antibodies. Cut-off point(s): any staining in the squamous cells	Describe reference test: HPV 16 ISH for cytology: FFPE cell blocks. Signal amplification for visualization of HPV16 infected cells. HPV16 for histology: 5-micron sections were deparaffinized. Slides were hybridized with HPV16 probes (and in specific cases with 6, 11, 18, 31, 33, 35, 45, 51, 52). Cut-off point(s): hybridization signals visualized as dots in squamous cell nuclei.	Time between the index test en reference test: Unclear For how many participants were no complete outcome data available? 9 cases did not undergo reference testing when cytology was compared to histology. Reasons for incomplete outcome data described. NR	Outcome measures and effect size (include 95%CI and p-value if available): <u>Oropharynx only (n=19), FNA material, p16 versus HPV16 ISH:</u> TP: 9 FP: 4 FN: 1 TN: 5 Sensitivity: 0.90 (0.55-1.00) Specificity: 0.56 (0.21-0.86) *sens/spec calculated from presented data. <u>Oropharynx only (n=10), p16 (FNA material) versus</u>	Because a database was searched and FNA smeas had to be available as well as the surgical specimen, a selection of patients might occur. It is possible that not all patients suspected of HPV positivity presenting in the hospital are included in the sample.

						<u>HPV16 ISH (resection material):</u> TP: 0 FP: 4 FN: 2 TN: 4 9 cases did not undergo reference testing and were excluded. Sensitivity: 0.00 (0.00-0.84) Specificity: 0.50 (0.16-0.84) *sens/spec calculated from presented data. <u>CUP only (n=10), FNA material, p16 versus HPV16 ISH:</u> TP: 3 FP: 0 FN: 0 TN: 7 Sensitivity: 1.00 (0.29-1.00) Specificity: 1.00 (0.59-1.00) *sens/spec calculated from presented data.	

Bishop 2012	Type of study: Consecutive Setting and country: Hospital, USA Funding and conflicts of interest: no financial disclosures. Funded by NIDCR.	Inclusion criteria: not reported. Exclusion criteria: Not reported N= 24 (27 cytologic preparations), n=12 with a lymph node sample site Prevalence: 5/12 were HPV positive by HPV ISH Mean age ± SD: Not reported Sex: not reported Tumor site (n=12): Unknown: 1 Skin: 2 Larynx: 2 Floor of mouth: 1 Tongue: 1 Tongue base: 1 Tonsil: 4	A 12-gauge needle was used for aspirates. Multiple passes were made. Hybrid Capture 2: Detects HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68. Specimen DNA was denatured into single stranded DNA and hybridized with specific probes. RNA/DNA hybrids are immobilized onto microplate surface. Light is emitted and relative light units are measured. The intensity of the light denotes the presence or absence of HPV DNA. Cut-off point(s): HC2: ≥3 RLU/CO was positive, 0.85-3 was equivocal, <0.85 negative.	Describe reference test: HPV ISH. Hybridization was performed with the HPV III Family 16 probe set and captures HPV 16, 18, 33, 35, 45, 51, 52, 56, 66. 5-micron sections from tissue microarrays and FFPE tumor blocks were used. Cut-off point(s): Hybridization signals localized to tumor cell nuclei defined an HPV positive tumor.	Time between the index test en reference test: Not reported For how many participants were no complete outcome data available? NA Reasons for incomplete outcome data described. NA	Outcome measures and effect size (include 95%CI and p-value if available): <u>HC2 versus HPV ISH, lymph node sample site only (n=12):</u> TP: 5 FP: 0 FN: 0 TN: 7 Sensitivity: 1.00 (0.48-1.00) Specificity: 1.00 (0.59-1.00) *sens/spec calculated from presented data.	P16 was performed on resected HNSCCs. Cytologic material from excised HNSCC of known HPV status
Buonocore 2019	Type of study: Consecutive Setting and country: Hospital, USA Funding and conflicts of	Inclusion criteria: previous or unknown oropharyngeal HNSCC (or determined to have HNSCC at the time of procedure), unknown p16 status.	FNA typically under ultrasound guidance with a 25-gauge needle. Initial passes were performed for Diff-Quik-staned and ethanol fixed smears. Multiple passes were made and allowed to clot before a transfer into	Describe reference test: HPV ISH targeting HPV 16, 18, 26, 31, 33, 35, 39, 41, 45, 51, 52, 53, 56, 58, 59, 66, 68, 73, 82 and E6/E7 mRNA. Unclear on which material ISH was performed.	Time between the index test en reference test: Unclear For how many participants were no complete	Outcome measures and effect size (include 95%CI and p-value if available): <u>P16 versus HPV ISH, Formalin-</u>	

	interest: authors had no COI to disclose, funding was supported by cancer center support grant for the national institutes of health/national cancer center institute	Exclusion criteria: Not reported. N= 25 Prevalence: 24/24 positives for mRNA ISH (1 case was nonconstributory) Mean age (range): 60 (47-76) Sex: 22M / 3F Smoking history: Never: 14 History 11 (pack-years range: 0.5-40, one person smoked 30 cigars per year) Alcohol use: Never: 1 Abstinent: 1 (9 year abstinent) Occasional: 2 Social: 15 Daily: 3 Heavy: 3	formalin fixation. From this material a CytoLyt and a formalin-fixed cellblock were made. Describe index test: P16 CytoLyt cellblock: Multiple passes were made and allowed to clot before a transfer into formalin fixation. From this material a CytoLyt and a formalin-fixed cellblock were made. E6H4 mouse antihuman monoclonal antibodies were used. P16 formalin-fixed cellblock: Multiple passes were made and allowed to clot before a transfer into formalin fixation. From this material a CytoLyt and a formalin-fixed cellblock were made. E6H4 mouse antihuman monoclonal antibodies were used. Cut-off point(s): ≥70% of the tumor cells showing nucleas and cytoplasmic staining.	Cut-off point(s): Not described how a positive case was defined.	outcome data available? 1/25 Reasons for incomplete outcome data described. Insufficient tumor remaining on deeper levels.	<u>fixed cell blocks (n=24):</u> TP: 23 FP: 0 FN: 1 TN: 0 Sensitivity: 0.96 (0.79-1.00) Specificity: Not Estimable. (0 TN) One case had a non-contributory reference test and was excluded from analysis. *sens/spec calculated from presented data. <u>P16 versus HPV ISH, CytoLyt cell blocks (n=24):</u> TP: 4 FP: 0 FN: 20 TN: 0 Sensitivity: 0.17 (0.05-0.37) Specificity: Not Estimable. (0 TN) One case had a non-contributory reference test and was excluded from analysis.	

						*sens/spec calculated from presented data	
Chute 2014	Type of study: Prospective Setting and country: Hospital, USA Funding and conflicts of interest: authors had no COI to disclose, funding was supported by the American society of cytopathology foundation through an ASC cytology research seed grant.	Inclusion criteria: HN-site FNA sample interpreted as SCC / atypical/ suspicious for SCC. Exclusion criteria: Not reported N=95 (96 FNAs) Prevalence: 29/95, 30.5% Median age (range): 60 (17-93) Sex: 72M / 21F Primary tumor location, n: Orophatynx: 32 Non-oropharyngeal: 32 Non-head/neck: 18 Unknown: 13	FNA Describe index test: CISH for HR HPV: The Inform HPV III Family 16 probe was used, which recognizes HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58. HC2 was performed according to the manufacturer's instructions. HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68 are detected. Microplate surfaces are coated with antibodies. Light is emitted and measured. Cut-off point(s): CISH: discrete dotlike blue staining in tumor nuclei was defines as a positive case. HC2: ≥1 RLU/CO was defined as a positive case.	FFPE block of primary tumor or excised neck metastasis were used when available. Describe reference test: P16 IHC was performed using the i-View DAB streptavidin-biotin-based detection kit. CISH: The Inform HPV III Family 16 probe was used, which recognizes HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58. Cut-off point(s): Positive results from both p16 and CISH defined a positive case. CISH: discrete dotlike blue staining in tumor nuclei was defines as a positive case. P16: >75% strong and diffuse tumor cytoplasmic and nuclear staining	Time between the index test en reference test: Not reported For how many participants were no complete outcome data available? 19 missings for HC2 cyto versus surgical 44 missings for CISH cyto versus surgical Reasons for incomplete outcome data described? samples were not included for analyses when there were equivocal results, no surgical hpv status could be determined or there were acellular cellblocks	Outcome measures and effect size (include 95%CI and p-value if available): <u>HC2 cytology versus surgical specimen (p16+CISH) (n=77)</u> TP: 21 FP: 0 FN: 8 TN: 48 Sensitivity: 0.72 (0.53-0.87) Specificity: 1.00 (0.93-1.00) 15 cases were not tested. 4 cases had equivocal outcomes on the reference test (1 tested positive, 3 tested negative on HC2) *sens/spec calculated from presented data.	Cytology and surgical specimen analyses were interpreted blinded to each other's results.

						<u>CISH cytology versus surgical specimen (p16+CISH) (n=52)</u> TP: 11 FP: 7 FN: 7 TN: 27 Sensitivity: 0.61 (0.36-0.83) Specificity: 0.79 (0.62-0.91) 7 cases were not tested (from the table), however 44 cases were not tested compared to the overall included sample. *sens/spec calculated from presented data. <u>HC2 cytology versus surgical specimen (p16+CISH) on the same sample (n=52) as tested with CISH cyto:</u> TP: 17 FP: 0 FN: 1 TN: 34	

						Sensitivity: 0.94 (0.73-1.00) Specificity: 1.00 (0.90-1.00) 44 cases were not tested compared to the overall included sample. *sens/spec calculated from presented data	
Hou 2016	Type of study: Database Setting and country: Hospital, USA Funding and conflicts of interest: authors had no COI to disclose. Funding was not reported in the manuscript	Inclusion criteria: Metastatic HNSCC in cervical lymph nodes diagnosed by FNA, p16 and HPV ISH was performed on FNA material. Exclusion criteria: Not reported. N=87 Prevalence: 42/87 Mean age (range): 59 (38-86) Sex: 80M / 7F Other important characteristics: Primary site, n: Base of tongue: 32	FNA specimen were centrifuged. The specimen was clot dried and places in a CellSafe mesh calsule. The capsule was fixated in formalin (10% neutral buffered). Describe index test: P16 on cell block sections. Monoclonal antibodies were used (E6H4). Cut-off point(s): ≥ 70% diffuse or strong staining in nuclei and cytoplasm	Test on FNA material. Describe reference test: HPV ISH was performed according to the manufacturer's protocol on 4-micrometer microtome sections (FNA cell block). HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, code Y1443). Cut-off point(s): Presence of staining in nuclei	Time between the index test en reference test: Not reported For how many participants were no complete outcome data available? 14 (14/87) were excluded form analyses because specimens were deemed inadequate. Reasons for incomplete outcome data described? Specimens were deemed inadequate	Outcome measures and effect size (include 95%CI and p-value if available): <u>P16 versus HPV ISH on FNA material (n=73)</u> TP: 41 FP: 15 FN: 1 TN: 16 Sensitivity: 0.98 (0.87-0.1.00) Specificity: 0.52 (0.33-0.70 PPV: 0.73 NPV: 0.94 14 cases were not tested compared to the overall included sample	Because a database was searched, both p16 and ISH had to be performed on FNA material. A selection of patients might occur. It is possible that not all patients suspected of HPV positivity presenting in the hospital are included in the sample.

		Tonsil: 19 Other oropharyngeal: 4 Non oropharyngeal: 25 Unknown: 7				(inadequate specimen).	
Jalaly 2015	Type of study: Database Setting and country: Hospital, USA Funding and conflicts of interest: two authors are employees of Advanced Cell Diagnostics and own stock in the company. One author reported to receive research support from Advanced Cell Diagnostics outside the submitted work. The work was funded by the Washington University Department of Pathology and Immunology. ISH was performed and	Inclusion criteria: FNA diagnosis for metastatic SCC to cervical lymph nodes, clinically proven SCC with corresponding surgical specimen (either biopsy or resection) Exclusion criteria: Cell blocks have no tumor cells. N= 48 Prevalence: 20/28 cases were positive by ISH (n=20 did not undergo ISH) Mean age ± SD: 60.8 (11.7) Sex: 44M / 4F FNA site, n: Neck: 41 Subcarinal: 2 Mediastinal: 1 Submandibular: 2	FNA specimens were centrifuged for 2 minutes (x800g). The sediment was evaluated, and CytoRich Red was added with repeated centrifuging when the sediment was bloody. Isotonic saline was added when the sediment was sparse, or CytoRich Red was used. When the specimen was sparse, toluidine blue was added as well. O plasma was added. The specimen clot was dried on tissue paper in placed in a CellSafe calsule. The capsule was inserted into a cassette and fixed in formalin (10% neutral-buffered) Describe index test: P16 was performed in theFNA cell blocks using a monoclonal antibody (clone E6H4) following a standard protocol. Scoring of cell blocks were performed without the	ISH on recuts of FNA cell blocks from specimen that had ≤50% p16 staining that had available tumor cells and on specimens >50% p16 staining that had abundant cellularity on the recut. Describe reference test: ISH detected E6/E7 mRNA (RNAscope HPV kit). ISH was performed according to the manufacturer's instruction. 4-mm FFPE cellblock sections were prepared. Cases were excluded when ubiquitin C signal was less or equal to Dap B signal. Staining was scored independently by 2 pathologists. Cut-off point(s): Red, punctate dots present in the nucleus and/or the cytoplasm. Positive cases had to have granular cytoplasmin and/or nuclear red staining that	Time between the index test en reference test: Unclear. For how many participants were no complete outcome data available? 20 (20/48) cases did not undergo the reference test. Reasons for incomplete outcome data described. ISH was not performed on 11 cases; Specimens we acellular for ISH in 6 cases; Specimens failed the quality control in 3 cases (ubiquitin C signal was less or equal to Dap B signal on ISH); Therefore, 20 cases were	Outcome measures and effect size (include 95%CI and p-value if available): <u>P16 (FNA) versus ISH (FNA), n=28:</u> TP: 6 FP: 0 FN: 14 TN: 8 Sensitivity: 0.30 (0.12-0.54) Specificity: 1.00 (0.63-1.00) ISH was not performed on 11 cases; Specimens we acellular for ISH in 6 cases; Specimens failed the quality control in 3 cases; Therefore, 20 cases were excluded from analysis.	Because a database was searched it is possible that not all patients suspected of HPV positivity presenting in the hospital are included in the sample. Reference testing (ISH) was performed on 28/48 of the included participants. P16 on cell blocks versus p16 on tissue was not extracted, since the reference test had to be a test that detected DNA/mRNA.

	funded by Advanced Cell Diagnostics.	Chest wall: 1 Supraclavicular: 1 Primary site: Base of tongue: 14 Tonsil: 15 Other oropharyngeal: 8 Oral cavity: 6 Larynx: 1 Maxilla: 1 Unknown: 3 Type of specimen: Biopsy: 16 Resection 32	knowledge of the p16 status from tissue. Cut-off point(s): P16: more than 70% nuclear and cytoplasmic staining of tumor cells	was higher than the signal on the DapB-negative control slide.	excluded from analysis.	*sens/spec calculated from presented data.	
Jannapureddy 2010	Type of study: Database Setting and country: Hospital, USA Funding and conflicts of interest: funding and CoI were not reported in the mansucript.	Inclusion criteria: Cell block cytologic diagnosis of metastatic SCC in cervical lymph nodes, between sept 2004-sept 2008 Exclusion criteria: inadequate cell block material N=45 (n=40 selected, because 5 had inadequate cell block material) Prevalence: 9/40 (22.5%)	FNA were fixed in 10% neutral-buffered formalin and centrifuged (7 min) Supernatant was discharged and the resulting content (collidon bag and contents) were submitted for routine histology. Cell block tissue (5 micrometer) of FFPE tissue were tested. Describe index test: P16: monoclonal p16(INK4a) (E6H4) incubation was performed for 30 minutes at room temperature.	Cell block tissue (5 micrometer) of FFPE tissue were tested. Describe reference test: HPV ISH detected HPV 16, 18, 31, 33, 51. Cell block tissue was deparaffinized and rehydrated. Slides were air-dried after rinsing. For 5 minutes, sections were denatured (95 deg. Celsius). Sections were hybridized overnight at 37 deg. Celsius. Sections were incubated. Cut-off point(s):	Time between the index test en reference test: Unclear For how many participants were no complete outcome data available? 5 cases were excluded from selection because they had inadequate cell blocks. Reasons for incomplete outcome data described?	Outcome measures and effect size (include 95%CI and p-value if available): <u>P16 (FNA) versus ISH (FNA), n=40:</u> TP: 9 FP: 7 FN: 0 TN: 24 Sensitivity: 1.00 (0.66-0.1.00) Specificity: 0.77 (0.59-0.90) 5 cases were excluded from patient selection because they had	Because a database was searched it is possible that not all patients suspected of HPV positivity presenting in the hospital are included in the sample. Participants needed to have had a cytologic diagnosis of metastatic SCC in cervical lymph nodes.

		Mean age: 58.2 (range: 25-87) Sex: 36M / 4F Primary tumor site, n: Oropharyngeal: 11 Nasopharyngeal: 2 Other: 5 Not determined: 9	ProExC: incubation was performed for 30 minutes at room temperature. Cut-off point(s): P16: nuclear and cytoplasmic staining ProExC: nuclear staining.	Positive cases were defined as punctuate or dot-like nuclear staining.	inadequate cell blocks	inadequate cell blocks, therefore the analyses were performed on n=40 *sens/spec calculated from presented data. <u>ProExC (FNA) versus ISH (FNA), n=40:</u> TP: 9 FP: 26 FN: 0 TN: 5 Sensitivity: 1.00 (0.66-0.1.00) Specificity: 0.16 (0.05-0.34) 5 cases were excluded from selection because they had inadequate cell blocks, therefore the analyses were performed on n=40 *sens/spec calculated from presented data	
Sivars 2017	Type of study: Prospective	Inclusion criteria: Suspected of HNSCC or with neck masses	DNA was extracted from FNA cytology samples. QIAmp DNA micro kit was	Some cases had FFPE biopsies from a primary tumor.	Time between the index test en reference test:	Outcome measures and effect size (include 95%CI	Patients included in the sample with malignant conditions othat

	Setting and country: Hospital, Sweden Funding and conflicts of interest: COI were not declared in the manuscript. Contract grant sponsor was the Swedish Cancer Society / The Stockholm Cancer Society, Ther cancer and Allergy foundation, Karolinska Institutet, The Stockholm City council	suspicious for metastasis. Exclusion criteria: Not enough material left for testing after cytological diagnosis. N=66 Prevalence: 17/66 (23 cases not tested) Mean age ± SD: 61.0 (oropharyngeal, n=20) 71.5 (other malignancies, n=17) 53.0 (benign conditions, n=29) Sex: 35M / 31F FNAC sites, n: Primary tumor: 2 Neck masses: 64	used to extract DNA from thawed aspirates. Describe index test: FNAC material; HPV DNA multiplex assay (lab): 27 HPV types were tested using GP5+/GP6+ primers. Additionally, E6 (HPV 16, 33) were amplified. DNA detection was performed on bead-based multiplex was used. Mean fluorescent intensity was used. HPV RT-PCR (clinic): FNAC material; a real-time TaqMan PCR was performed to detect the 7 most common HR-HPV genotypes. Cut-off point(s): HPV DNA multiplex assay: mean fluorescent intensity above background*1.5+15 was considered positive for HPV. HPV PCR: unclear	Describe reference test: HPV DNA multiplex assay (histo): 27 HPV types were tested using GP5+/GP6+ primers. Additionally, E6 (HPV 16, 33) were amplified. DNA detection was performed on bead-based multiplex was used. Mean fluorescent intensity was used. HPV RT-PCR (histo): tested on FFPE material. a real-time TaqMan PCR was performed to detect the 7 most common HR-HPV genotypes. Cut-off point(s): mean fluorescent intensity above background*1.5+15 was considered positive for HPV.	For how many participants were no complete outcome data available? For comparison on histopathologic tissue 46 cases were excluded (no histopathology other than p16 was performed) Reasons for incomplete outcome data described? Unclear, some were not tested, some were not possible to test, some were not available	and p-value if available): <u>HPV RT-PCR at clinic (cyto) versus HPV multiplex assay (cyto), n=40:</u> TP: 15 FP: 0 FN: 0 TN: 25 Sensitivity: 1.00 (0.78-0.1.00) Specificity: 1.00 (0.86-1.00) 26 cases were not tested with the HPV RT-PCR *sens/spec calculated from presented data. <u>HPV RT-PCR at clinic (cyto) versus HPV multiplex assay (histo FFPE), n=9:</u> TP: 8 FP:0 FN: 0 TN: 1 Sensitivity: 1.00 (0.63-0.1.00) Specificity: 1.00 (0.03-1.00) 56 cases were not tested on	than oropharyngeal SCC or with benign conditions did not undergo HPV multiplex assay and HPV PCR on histopathologic FFPE tissue. Results from HPV mRNA detection were not extracted, since this was only tested on HPV positive specimen.

						histopathology with HPV multiplex, 1 case was not tested with HPV RT-PCR. Therefore 57 cases were excluded from the analysis. *sens/spec calculated from presented data. <u>HPV multiplex assay (cyto) versus HPV multiplex assay (histo FFPE), n=10:</u> TP: 9 FP: 0 FN: 0 TN: 1 Sensitivity: 1.00 (0.66-1.00) Specificity: 1.00 (0.03-1.00) 56 cases were not tested on histopathology with HPV multiplex; therefore 56 cases were excluded from the analysis.	

						*sens/spec calculated from presented data	
Smith 2014	Type of study: Prospective Setting and country: Hospital, USA Funding and conflicts of interest: Authors declared there were no COI. Funded by an NIH/NCI Grant	Inclusion criteria: Cervical lymph nodes ≥1cm, Exclusion criteria: Not reported N=25 Prevalence: 9 cases did not receive HPV ISH, 11/16 were HPV-positive Mean age ± SD: NR. Sex: NR Cause of cervical lymph mass: Tumor/matastasis: 18 Lymphadenopathy: 2 Non-inflammatory process: 5 Tumor location, n: Palatine tonsils: 6 Base of tongue: 8 Unknown: 2 Hypopharynx: 1 Skin of the auricle: 1	FNA biopsies were performed of metastatic cervical lymph nodes (in the clinic or operating room). 25-gauge needle was used passing into the lymph node 3-5 times. The aspirate was places on a slide (charged with Vista Vision HistoBond). The slide was air-dried and stained with a Diff-Quik stain. A final pass was made with a fresh needle into the lymph node. This aspirate was placed into 1ml of transport medium and stored at -20 deg. Celsius until HC2 was performed. Describe index test: A modified-HC2 HPV assay: The test detects HPV 16, 18, 31, 33, 35, 39, 45, 51, 52,56, 58, 59, 68. DNA was denatured to single standed DNA and incubated in 65 deg. Celsius water for 45 minutes. Samples were applied to microplate wells coaded with hybrid specific antibodies. Signal	Reference test was performed on tissue specimen from patients who had a resection. 5-micron FFPE tissue sections from tumors/biopsies were conditioned. Describe reference test: ISH hybridization was performed with HPV III Family16 probe set (HPV 16, 18, 31, 33, 35, 39, 45, 51,52, 56, 58, 66)/ Genotypes were detected with the ISH iVIEW Blue Plus Detection Kit. Cut-off point(s): Punctate signals in the nuclei defined a positive case.	Time between the index test en reference test: Unclear For how many participants were no complete outcome data available? N (%) Reasons for incomplete outcome data described?	Outcome measures and effect size (include 95%CI and p-value if available): <u>HC2 (cyto) versus HPV ISH (histo FFPE), n=15:</u> TP: 7 FP: 0 FN: 3 TN: 5 Sensitivity: 0.70 (0.35-0.93) Specificity: 1.00 (0.48-1.00) 9 cases did not undergo the ISH reference testing; 1 case had an equivocal result for HC2 and was excluded from analysis. Therefore 10 cases were excluded from analysis. *sens/spec calculated from presented data.	Data for p16 was not extracted, since p16 was performed on tissue specimens from the resection. The reference test had to be a test that detects HPV DNA/mRNA.

			amplification was performed with Detection Reagent II. Light emission was used to detect HPV DNA Cut-off point(s): A positive case was defined as RLU/CO ≥2.5. A negative case was defined as RLU/CO <0.85. Between 0.85-2.5 was considered equivocal.				
Takes 2016	Type of study: Database Setting and country: Hospital, the Netherlands Funding and conflicts of interest: Authors declared there were no COI. No funding received.	Inclusion criteria: FFPE histological material from the primary tumor or cetastatic lymph node available, FFPE material was both hpv AND p16 negative or both hpv AND p16 positive. Exclusion criteria: Secondary tumor in the head/neck region, insufficient cytological material, previous exposure of the neck to radiotherapy. N=47 Prevalence: 25/47	Scraped FNAC from archival slided were used and DNA was purified. DNA was diluted in elution buffer and stored at -20 deg. Celsius until processing by PCR. Describe index test: HPV PCR testing combined (cyto): p16 procedures were not described. For the HPV PCR, a broad-spectrum DNA amplification was performed using a short-PCR-fragment assay. 9 conservative probes were used in a micro titer hybridization assay (DNA enzyme immunoassay, DEIA). Cases positive for HPV by DEIA were analysed with line-specific	DNA was isolated from FFPE tissue sections (4 micrometer) and used for PCR analysis. Describe reference test: P16 + HPV PCR testing combined (histo): p16 procedures were not described. For the HPV PCR, a broad-spectrum DNA amplification was performed using a short-PCR-fragment assay. 9 conservative probes were ised in a micro titer hybridization assay (DNA enzyme immunoassay, DEIA). Cases positive for HPV by DEIA were analysed with line-specific probes assay (LiPA25) for HPV 6, 11, 16, 18, 31, 33, 34,35, 39, 40, 42, 43, 44,	Time between the index test en reference test: Not reported For how many participants were no complete outcome data available? None Reasons for incomplete outcome data described. NA	Outcome measures and effect size (include 95%CI and p-value if available): <u>HPV PCR (cyto) versus HPV PCR +p16 (histo FFPE), n=47:</u> TP: 24 FP: 0 FN: 1 TN: 22 Sensitivity: 0.96 (0.80-1.00) Specificity: 1.00 (0.85-1.00) *sens/spec calculated from presented data.	Because a database was searched it is possible that not all patients suspected of HPV positivity presenting in the hospital are included in the sample. Specimen were selected only when FFPE material were both hpv and p16 positive or negative.

		Mean age: 58 (range: 28.9-77.2) Sex: 33M / 14F T-stage, n: T0: 6 T1: 7 T2: 16 T3: 11 T4: 7 N-stage, n: N0: 2 N1: 7 N2: 38 N3: 0 Primary tumor site, n: Tonsils: 19 Base of tongue: 12 Unknown: 6 Other oropharyngeal: 10 Origin of FFPE material: Tonsils: 21 Base of tongue: 9 Neck dissection: 7 Other oropharyngeal: 10 Histological tumor differentiation, n: Poor: 7 Moderate: 33	probes assay (LiPA25) for HPV 6, 11, 16, 18, 31, 33, 34,35, 39, 40, 42, 43, 44, 45, 51, 52, 53, 54, 56, 58, 59, 66, 68/73, 70, 74. LiPA strips were visually inspected and interpreted following the standardized reference guide. Cut-off point(s): Unclear	45, 51, 52, 53, 54, 56, 58, 59, 66, 68.73, 70, 74. LiPA strips were visually inspected and interpreted dollowing the standardized reference guide. Cut-off point(s): Unclear (both p16 and HPV PCR)			

		Well: 3 Unknown: 4					
Xu 2016	Type of study: Database Setting and country: Hospital, USA Funding and conflicts of interest: Authors declared that there were no CoI. Partly supported by Cancer center support grant of the national institutes of health / national cancer institute.	Inclusion criteria: Metastatic SCC of a cervical lymph node based of FNA between 2007 and 2015, p16 performed on both cytological material and corresponding surgical material. Exclusion criteria: Tumors originating from outside head and neck region. N = 60 (63 FNAs from neck lymph nodes, 66 surgical specimens (47 primary tumor, 19 metastases)) Prevalence: 18/30 in cytology p16 versus CISH Mean age ± SD: Not reported Sex: Not reported Tumor location, n: Non-oropharyngeal: 6 Oropharyngeal: 48	FNA specimens were collected for cell blocks in ThinPrep and CytoLyt solution. Direct smears were fixed using ethanol (95%) Describe index test: P16 (cyto) was performed on cell block, smear, or ThinPrep. Cut-off point(s): Several thresholds, P16 nuclear and cytoplasmic staining in 1%/ 5%/ 10%/ 15%/ 70%.	CISH was performed on cytologic material. Describe reference test: CISH (cyto) was performed using high risk HPV probes (HPV 16, 18, 31, 33, 51). Cut-off point(s): CISH was categorized as negative, positive or equivocal. Negative and equivocal cases were not defined. A positive case was defined as discrete dot-like stippled nuclear labelling.	Time between the index test en reference test: For how many participants were no complete outcome data available? Overall, 8 FNAs and 53 surgical specimenswere excluded from analysis (p16 versus CISH) Reasons for incomplete outcome data described. 8 FNAs were excluded because the percentage p16 was not documented and/or cases were interpreted as equivocal in CISH. 53 surgical specimen were excluded, since the presented data concerned testing on cytology only. (p16 on FNA was	Outcome measures and effect size (include 95%CI and p-value if available): <u>P16 (cyto) versus HPV ISH (cyto), n=30, at 1% cut-off:</u> TP: 18 FP: 4 FN: 0 TN: 8 Sensitivity: 1.00 (0.81-1.00) Specificity: 0.67 (0.35-0.90) PPV: 0.82 NPV: 1.00 8 cases did not have their p16 status documented and/or were interpreted as equivocal by CISH. *sens/spec calculated from presented data. <u>P16 (cyto) versus HPV ISH (cyto),</u>	Because a database was searched it is possible that not all patients suspected of HPV positivity presenting in the hospital are included in the sample. Only cases with p16 performed on both the cytological and corresponding surgical materials were included. CISH was performed in 38 cytological and 53 surgical samples. Data for p16 on the surgical specimen were not extracted, since the reference test should be a test that detects HPV DNA/mRNA. P16 on cytological material was performed on 3 different types of preparations (cell

		Unknwon primary: 6			not related to CISH on surgical specimens).	<u>n=30, at 5% cut-off:</u> TP: 17 FP: 4 FN: 1 TN: 8 Sensitivity: 0.94 (0.73-1.00) Specificity: 0.67 (0.35-0.90) PPV: 0.81 NPV: 0.89 8 cases did not have their p16 status documented and/or were interpreted as equivocal by CISH. *sens/spec calculated from presented data. <u>P16 (cyto) versus HPV ISH (cyto), n=30, at 10% cut-off:</u> TP: 17 FP: 3 FN: 1 TN: 9 Sensitivity: 0.94 (0.73-1.00) Specificity: 0.75 (0.43-0.95) PPV: 0.85	block, direct smear, ThinPrep). Sub-analyses are not reported for each preparation (and the data is not reported to perform such subanalyses).

						NPV: 0.90 8 cases did not have their p16 status documented and/or were interpreted as equivocal by CISH. *sens/spec calculated from presented data. <u>P16 (cyto) versus HPV ISH (cyto), n=30, at 15% cut-off:</u> TP: 17 FP: 3 FN: 1 TN: 9 Sensitivity: 0.94 (0.73-1.00) Specificity: 0.75 (0.43-0.95) PPV: 0.85 NPV: 0.90 8 cases did not have their p16 status documented and/or were interpreted as equivocal by CISH.	

						*sens/spec calculated from presented data. <u>P16 (cyto) versus HPV ISH (cyto), n=30, at 70% cut-off:</u> TP: 9 FP: 0 FN: 9 TN: 12 Sensitivity: 0.50 (0.26-0.74) Specificity: 1.00 (0.74-1.00) PPV: 0.60 NPV: 0.58 8 cases did not have their p16 status documented and/or were interpreted as equivocal by CISH. *sens/spec calculated from presented data	

[1] In geval van een case-control design moeten de patiëntkarakteristieken per groep (cases en controls) worden uitgewerkt. NB; case control studies zullen de accuratesse overschatten (Lijmer,1999).

[1] De referentiestandaard is de test waarmee definitief wordt aangetoond of iemand al dan niet ziek is. Idealiter is de referentiestandaard de Gouden standaard (100% sensitief en 100% specifiek). Let op! dit is niet de "comparison test/index 2".

[4] Beschrijf de statistische parameters voor de vergelijking van de indextest(en) met de referentietest, en voor de vergelijking tussen de indextesten onderling (als er twee of meer indextesten worden vergeleken).

Risk of bias assessment diagnostic accuracy studies (QUADAS II, 2011)

Study reference	Patient selection	Index test	Reference standard	Flow and timing	Comments with respect to applicability
Baldassari 2015	Was a consecutive or random sample of patients enrolled? Unclear, reason: data was collected prospectively but it was not reported this was done consecutively or randomly. Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Unclear, reason: in and exclusion criteria were not stated.	Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? Unclear	Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Unclear	Was there an appropriate interval between index test(s) and reference standard? Unclear Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? No, reason atypical and suspicious cytological results were excluded from analysis	Are there concerns that the included patients do not match the review question? Unclear Are there concerns that the index test, its conduct, or interpretation differ from the review question? No Are there concerns that the target condition as defined by the reference standard does not match the review question? Unclear, no threshold for the reference standard was provided.
	CONCLUSION: Could the selection of patients have introduced bias? **RISK: UNCLEAR**	CONCLUSION: Could the conduct or interpretation of the index test have introduced bias? **RISK: UNCLEAR**	CONCLUSION: Could the reference standard, its conduct, or its interpretation have introduced bias? **RISK: UNCLEAR**	CONCLUSION Could the patient flow have introduced bias? **RISK: HIGH**	
Begum 2007	Was a consecutive or random sample of patients enrolled?	Were the index test results interpreted without knowledge of	Is the reference standard likely to correctly classify the target condition?	Was there an appropriate interval between index test(s)	Are there concerns that the included patients do not match the review question? Unclear

	No, reason: a database was searched. Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Unclear, exclusion criteria were not reported.	the results of the reference standard? Yes If a threshold was used, was it pre-specified? Yes	Yes Were the reference standard results interpreted without knowledge of the results of the index test? Unclear	and reference standard? Unclear Did all patients receive a reference standard? No, reason: 9 cases did not receive the reference standard in some analyses. Did patients receive the same reference standard? Yes Were all patients included in the analysis? No	Are there concerns that the index test, its conduct, or interpretation differ from the review question? No Are there concerns that the target condition as defined by the reference standard does not match the review question? No
	CONCLUSION: Could the selection of patients have introduced bias? **RISK: HIGH**	CONCLUSION: Could the conduct or interpretation of the index test have introduced bias? **RISK: LOW**	CONCLUSION: Could the reference standard, its conduct, or its interpretation have introduced bias? **RISK: UNCLEAR**	CONCLUSION Could the patient flow have introduced bias? **RISK: HIGH**	
Bishop 2012	Was a consecutive or random sample of patients enrolled? Yes, consecutive. Was a case-control design avoided? Yes	Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? Yes	Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the	Was there an appropriate interval between index test(s) and reference standard? Unclear Did all patients receive a reference standard? Yes	Are there concerns that the included patients do not match the review question? Unclear Are there concerns that the index test, its conduct, or interpretation differ from the review question? No

	Did the study avoid inappropriate exclusions? Unclear, in and exclusion criteria were not described.		results of the index test? Unclear	Did patients receive the same reference standard? Yes Were all patients included in the analysis? Yes	Are there concerns that the target condition as defined by the reference standard does not match the review question? No
	CONCLUSION: Could the selection of patients have introduced bias? **RISK: LOW**	CONCLUSION: Could the conduct or interpretation of the index test have introduced bias? **RISK: UNCLEAR**	CONCLUSION: Could the reference standard, its conduct, or its interpretation have introduced bias? **RISK: UNCLEAR**	CONCLUSION Could the patient flow have introduced bias? **RISK: LOW**	
Buonocore 2019	Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Unclear, no exclusion criteria reported.	Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? Yes	Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Unclear	Was there an appropriate interval between index test(s) and reference standard? Unclear Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis?	Are there concerns that the included patients do not match the review question? Unclear Are there concerns that the index test, its conduct, or interpretation differ from the review question? No Are there concerns that the target condition as defined by the reference standard does not match the review question? Unclear, no cut-off for reference was reported

				No, 1 non-contributory case was excluded. However, we judged that the exclusion of 1 case did not have a significant impact on the findings.	
	CONCLUSION: Could the selection of patients have introduced bias? **RISK: LOW**	CONCLUSION: Could the conduct or interpretation of the index test have introduced bias? **RISK: UNCLEAR**	CONCLUSION: Could the reference standard, its conduct, or its interpretation have introduced bias? **RISK: UNCLEAR**	CONCLUSION Could the patient flow have introduced bias? **RISK: LOW**	
Chute 2014	Was a consecutive or random sample of patients enrolled? Unclear, prospective data collection. It is not clear whether it was consecutive (or random). Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Unclear, no exclusion criteria are reported.	Were the index test results interpreted without knowledge of the results of the reference standard? Yes If a threshold was used, was it pre-specified? Yes	Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Yes	Was there an appropriate interval between index test(s) and reference standard? Unclear Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? No, cases were excluded when there were equivocal results, no surgical hpv status	Are there concerns that the included patients do not match the review question? Unclear Are there concerns that the index test, its conduct, or interpretation differ from the review question? No Are there concerns that the target condition as defined by the reference standard does not match the review question? No

				could be determined, or when there were acellular cell blocks.	
	CONCLUSION: Could the selection of patients have introduced bias? **RISK: UNCLEAR**	CONCLUSION: Could the conduct or interpretation of the index test have introduced bias? **RISK: LOW**	CONCLUSION: Could the reference standard, its conduct, or its interpretation have introduced bias? **RISK: LOW**	CONCLUSION Could the patient flow have introduced bias? **RISK: HIGH**	
Hou 2016	Was a consecutive or random sample of patients enrolled? No, a database was searched. Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Unclear, exclusion criteria were not reported.	Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? Yes	Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Unclear	Was there an appropriate interval between index test(s) and reference standard? Unclear Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? No, 14 cases were excluded because the specimen was deemed inadequate. It remains unclear what the influence of these exclusions are on the reported outcomes.	Are there concerns that the included patients do not match the review question? Unclear Are there concerns that the index test, its conduct, or interpretation differ from the review question? No Are there concerns that the target condition as defined by the reference standard does not match the review question? No

	CONCLUSION: Could the selection of patients have introduced bias? **RISK: HIGH**	CONCLUSION: Could the conduct or interpretation of the index test have introduced bias? **RISK: UNCLEAR**	CONCLUSION: Could the reference standard, its conduct, or its interpretation have introduced bias? **RISK: UNCLEAR**	CONCLUSION Could the patient flow have introduced bias? **RISK: UNCLEAR**	
Jalaly 2015	Was a consecutive or random sample of patients enrolled? No, a database was searched. Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? No, excluded patients could probably have a new pass to gather tumor cells by FNA when a prospective/consecutive design was used.	Were the index test results interpreted without knowledge of the results of the reference standard? Yes, but only blinded for p16 in tissue samples. Otherwise, it was unclear. If a threshold was used, was it pre-specified? Yes	Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Unclear	Was there an appropriate interval between index test(s) and reference standard? Unclear Did all patients receive a reference standard? No Did patients receive the same reference standard? Yes Were all patients included in the analysis? No, cases that received no reference test were excluded from analysis. Furthermore, acellular specimen and quality control-failed specimen were excluded as well	Are there concerns that the included patients do not match the review question? Unclear Are there concerns that the index test, its conduct, or interpretation differ from the review question? No Are there concerns that the target condition as defined by the reference standard does not match the review question? No
	CONCLUSION: Could the selection of patients have introduced bias?	CONCLUSION: Could the conduct or interpretation of the	CONCLUSION: Could the reference standard, its conduct,	CONCLUSION Could the patient flow have introduced bias?	

	RISK: HIGH	index test have introduced bias? **RISK: LOW**	or its interpretation have introduced bias? **RISK: UNCLEAR**	**RISK: HIGH**	
Jannapureddy 2010	Was a consecutive or random sample of patients enrolled? No, a database was searched. Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Unclear, no exclusion criteria were reported. However, 5 cases were not selected because they had inadequate cell block material.	Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? Yes	Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Unclear	Was there an appropriate interval between index test(s) and reference standard? Unclear Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? Yes	Are there concerns that the included patients do not match the review question? Unclear Are there concerns that the index test, its conduct, or interpretation differ from the review question? No Are there concerns that the target condition as defined by the reference standard does not match the review question? No
	CONCLUSION: Could the selection of patients have introduced bias? **RISK: HIGH**	CONCLUSION: Could the conduct or interpretation of the index test have introduced bias? **RISK: UNCLEAR**	CONCLUSION: Could the reference standard, its conduct, or its interpretation have introduced bias? **RISK: UNCLEAR**	CONCLUSION Could the patient flow have introduced bias? **RISK: LOW**	
Sivars 2017	Was a consecutive or random sample of patients enrolled? Unclear, not reported as such, however it may have been consecutive.	Were the index test results interpreted without knowledge of the results of the reference standard? Yes	Is the reference standard likely to correctly classify the target condition? Yes	Was there an appropriate interval between index test(s) and reference standard? Unclear	Are there concerns that the included patients do not match the review question? No

	Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes, probably.	If a threshold was used, was it pre-specified? No, not for HPV PCR on cytologic material.	Were the reference standard results interpreted without knowledge of the results of the index test? Yes	Did all patients receive a reference standard? No (56 cases were not tested with the HPV multiplex PCR on histologic material) Did patients receive the same reference standard? Yes Were all patients included in the analysis? No, 57 cases did not undergo the reference standard (HPV multiplex assay)	Are there concerns that the index test, its conduct, or interpretation differ from the review question? No Are there concerns that the target condition as defined by the reference standard does not match the review question? No
	CONCLUSION: Could the selection of patients have introduced bias? **RISK: LOW**	CONCLUSION: Could the conduct or interpretation of the index test have introduced bias? **RISK: HIGH**	CONCLUSION: Could the reference standard, its conduct, or its interpretation have introduced bias? **RISK: LOW**	CONCLUSION Could the patient flow have introduced bias? **RISK: HIGH**	
Smith 2014	Was a consecutive or random sample of patients enrolled? Unclear Was a case-control design avoided? Yes	Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? Yes	Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the	Was there an appropriate interval between index test(s) and reference standard? Unclear Did all patients receive a reference standard? No	Are there concerns that the included patients do not match the review question? Unclear Are there concerns that the index test, its conduct, or interpretation differ from the review question? No

	Did the study avoid inappropriate exclusions? Unclear, exclusion criteria were not reported.		results of the index test? Unclear	Did patients receive the same reference standard? Yes Were all patients included in the analysis? No, 9 cases did not receive a reference test and 1 case had equivocal results and was excluded in analysis	Are there concerns that the target condition as defined by the reference standard does not match the review question? No
	CONCLUSION: Could the selection of patients have introduced bias? **RISK: UNCLEAR**	CONCLUSION: Could the conduct or interpretation of the index test have introduced bias? **RISK: UNCLEAR**	CONCLUSION: Could the reference standard, its conduct, or its interpretation have introduced bias? **RISK: UNCLEAR**	CONCLUSION Could the patient flow have introduced bias? **RISK: HIGH**	
Takes 2016	Was a consecutive or random sample of patients enrolled? No, a database was searched. Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? No, cases were selected when both p16 and	Were the index test results interpreted without knowledge of the results of the reference standard? Unclear If a threshold was used, was it pre-specified? Unclear	Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Unclear	Was there an appropriate interval between index test(s) and reference standard? Unclear Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes	Are there concerns that the included patients do not match the review question? Yes Are there concerns that the index test, its conduct, or interpretation differ from the review question? Unclear Are there concerns that the target condition as defined by the reference standard does not match the review question?

	HPV DNA test were positive or negative. Therefore, cases with (for example) equivocal results or only 1 test positive/negative were excluded. This may not be representative for clinical practice.			Were all patients included in the analysis? Yes	Unclear
	CONCLUSION: Could the selection of patients have introduced bias? **RISK: HIGH**	CONCLUSION: Could the conduct or interpretation of the index test have introduced bias? **RISK: UNCLEAR**	CONCLUSION: Could the reference standard, its conduct, or its interpretation have introduced bias? **RISK: UNCLEAR**	CONCLUSION Could the patient flow have introduced bias? **RISK: LOW**	
Xu 2016	Was a consecutive or random sample of patients enrolled? No, a database was searched. Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Unclear, cases were selected when p16 was both performed on cytologic and resection material. Selected cases may or may not represent cases in clinical practice,	Were the index test results interpreted without knowledge of the results of the reference standard? Unclear, p16 status from cytology was blinded for p16 and dna status from tumors. However, the comparison of interest was both performed on cytologic material and it was unclear whether this was blinded. If a threshold was used, was it pre-specified? Yes	Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Yes	Was there an appropriate interval between index test(s) and reference standard? Unclear Did all patients receive a reference standard? Unclear: 8 cases were excluded because the p16 status was not documented Did patients receive the same reference standard? Unclear: 8 cases were excluded because the p16 status was not documented	Are there concerns that the included patients do not match the review question? Unclear Are there concerns that the index test, its conduct, or interpretation differ from the review question? No Are there concerns that the target condition as defined by the reference standard does not match the review question? No

	depending how common it was that both specimen was tested with p16.			Were all patients included in the analysis? No: 8 cases were excluded because the p16 status was not documented	
	CONCLUSION: Could the selection of patients have introduced bias? **RISK: HIGH**	CONCLUSION: Could the conduct or interpretation of the index test have introduced bias? **RISK: UNCLEAR**	CONCLUSION: Could the reference standard, its conduct, or its interpretation have introduced bias? **RISK: LOW**	CONCLUSION Could the patient flow have introduced bias? **RISK: UNCLEAR**	

Judgments on risk of bias are dependent on the research question: some items are more likely to introduce bias than others, and may be given more weight in the final conclusion on the overall risk of bias per domain:

Patient selection:

- **Consecutive or random sample has a low risk to introduce bias.**
- **A case control design is very likely to overestimate accuracy and thus introduce bias.**
- **Inappropriate exclusion is likely to introduce bias.**

Index test:

- **This item is similar to "blinding" in intervention studies. The potential for bias is related to the subjectivity of index test interpretation and the order of testing.**
- **Selecting the test threshold to optimise sensitivity and/or specificity may lead to overoptimistic estimates of test performance and introduce bias.**

Reference standard:

- **When the reference standard is not 100% sensitive and 100% specific, disagreements between the index test and reference standard may be incorrect, which increases the risk of bias.**
- **This item is similar to "blinding" in intervention studies. The potential for bias is related to the subjectivity of index test interpretation and the order of testing.**

Flow and timing:

- **If there is a delay or if treatment is started between index test and reference standard, misclassification may occur due to recovery or deterioration of the condition, which increases the risk of bias.**
- **If the results of the index test influence the decision on whether to perform the reference standard or which reference standard is used, estimated diagnostic accuracy may be biased.**
- **All patients who were recruited into the study should be included in the analysis, if not, the risk of bias is increased.**

Judgement on applicability:

Patient selection: there may be concerns regarding applicability if patients included in the study differ from those targeted by the review question, in terms of severity of the target condition, demographic features, presence of differential diagnosis or co-morbidity, setting of the study and previous testing protocols.
Index test: if index tests methods differ from those specified in the review question there may be concerns regarding applicability.
Reference standard: the reference standard may be free of bias but the target condition that it defines may differ from the target condition specified in the review question.

Quality assessment with AGREE-II (Brouwers, 2010) for the CAP Guideline (Lewis, 2018)

Domain	Item	Score (1 strongly disagree - 7 strongly agree)	Comments
SCOPE AND PURPOSE	The overall objective(s) of the guideline is (are) specifically described.	7	Explicitly stated
	The health question(s) covered by the guideline is (are) specifically described.	7	Six questions are explicitly stated
	The population (patients, public, etc.) to whom the guideline is meant to apply is specifically described.	6	Deducible, patients with oropharyngeal tumors
STAKEHOLDER INVOLVEMENT	The guideline development group includes individuals from all relevant professional groups	7	Members with molecular pathology, surgical, medical, radiation oncology experience were seated in the working group. A methodologist was also added.
	The views and preferences of the target population (patients, public, etc.) have been sought.	7	An advisory panel was created. Two patient representatives took seat.
	The target users of the guideline are clearly defined.	3	Deducible, however it does not seem to be explicitly stated.
RIGOUR OF DEVELOPMENT	Systematic methods were used to search for evidence.	7	Search and selection are described in the methods supplement
	The criteria for selecting the evidence are clearly described.	3	Selection criteria are described, but may not be reproducible for each search/key question.
	The strengths and limitations of the body of evidence are clearly described.	3	Risk of bias of included literature is described in supplemental tables 2-6, however it may not cover all relevant domains per study design.
	The methods for formulating the recommendations are clearly described.	5	It is described how the panel would develop recommendations, however no formal framework seems to be used (e.g. GRADE evidence to decision). The panel needed to identify the relevant evidence and make 4 key judgements.
	The health benefits, side effects, and risks have been considered in formulating the recommendations.	7	Harms and benefits were one of the key judgements the panel had to make. (supplement)
	There is an explicit link between the recommendations and the supporting evidence.	6	Not assessed for the full guideline. However, systematically searching and selecting evidence and making key judgements about this should ensure the link between recommendations
	The guideline has been externally reviewed by experts prior to its publication.	7	The guideline was peer-reviewed

	A procedure for updating the guideline is provided.	6	The guideline is updated every 4 years, or earlier when new evidence could alter the guideline recommendations. A specific procedure was not stated.
CLARITY OF PRESENTATION	The recommendations are specific and unambiguous. The different options for management of the condition or health issue are clearly presented	7	Seem specific. Includes a figure describing a complex algorithm.
	Key recommendations are easily identifiable.	7	Summarized in a box and clearly stated in the text.
APPLICABILITY	The guideline describes facilitators and barriers to its application.	4	Describes implementation of the statements/recommendations in practice, rather than the use/uptake of the guideline. (supplement)
	The guideline provides advice and/or tools on how the recommendations can be put into practice.	4	Describes implementation of the statements/recommendations in practice, rather than the use/uptake of the guideline. (supplement)
	The potential resource implications of applying the recommendations have been considered.	5	This has been discussed (supplement), but could have been more extensively reported.
	The guideline presents monitoring and/or auditing criteria.	1	There does not seem to be any monitoring/audit criteria.
EDITORIAL INDEPENDENCE	The views of the funding body have not influenced the content of the guideline.	4	Unclear. It was stated that the CAP provided funding for the project and that no industry funds were used. However, it is not explicitly stated that the CAP did not have any influence on the contents of the guideline.
	Competing interests of guideline development group members have been recorded and addressed.	6	Assessed and reported. Two participants were assessed to have potential conflicts of interest. It was not stated which actions followed the COI assessment.
OVERALL GUIDELINE ASSESSMENT	Rate the overall quality of this guideline.	5	Overall, it seems to be a methodologically transparent guideline. PICOs and in/exclusion criteria per question should probably have been reported for reproducibility/updating purposes.
	I would recommend this guideline for use. (yes / yes, with modifications / no)	Yes, with modifications	For example, clear statements of PICOs per question, clear statements of in/exclusion criteria per key question, extended RoB assessments for RCTs, more extensive evidence tables, GRADEing of evidence, use of an explicit/formal evidence to decision framework (e.g. GRADE evidence-to-decision)

Table of excluded studies

Author and year	Reason for exclusion
Allison 2015	conference abstract
Alison 2016	FNA (25 gauge) and core biopsies (20 gauge) were not analyzed separately.
Aramouni 2010	Probably a conference abstract from the American Society of Cytopathology 58th Annual Scientific Meeting Platform and Poster Presentations, Supplement to Cancer Cytopathology

Bernadt 2017	narratieve review
Bishop 2014	Editorial
Buckley 2018	Not a systematic review
Buonocore 2017	Conference abstract
Channir 2016	Conference abstract
Channir 2016	Reference test does not seem to be a DNA/mRNA test
Chernesky 2018	Oral and oropharyngeal tumors
Dictor 2011	sensitivity / specificity can not be calculated
Dona 2014	research question is phrased differently
El-Salem 2019	
Faquin 2014	Editorial
Faquin 2016	Conference abstract
Datima 2013	Conference abstract
Fatima 2012	Conference abstract
Fatima 2013	Conference abstract
Gargano 2019	Conference abstract
Gipson 2018	No diagnostic accuracy of algorithms
Griffith 2019	Editorial
Grimes 2013	Does not compare p16 immunocytochemistry to a reference standard
Guo 2012	conference abstract
Guo 2014	Unclear whether this assay is used in the Netherlands
Haeggblom 2017	No diagnostic accuracy of algorithms
Hakima 2015	Reference test does not seem to be a DNA/mRNA test
Han 2016	Unclear whether this assay is used in the Netherlands
Holmes 2015	No sub-analyses for NFA only (FNA cyto + FNA biopsy were analysed together)
Holmes 2014	Narrative review
Hou 2016	Conference abstract
Jalaly 2020	narative review
Jannapureddy 2010	sens/spec gegeven, ook na te rekenen. Let op: spec voor p16 klopt niet. Sample uit database met cervical SCC metastasis.
Jannapureddy 2009	Conference abstract
Jarboe 2012	Narrative review
Jensen 2018	No diagnostic accuracy of algorithms
Kerr 2014	FNA were taken from resections or biopsies. Only 9/33 were taken from lymph nodes.
Khazai 2017	Conference abstract
Krame 2013	Narrative review
Kwon 2018	Conference abstract
Lastra 2013	research question is phrased differently
Lewis 2018	tests on biopsies or resected tissue
Lewis 2018	Guideline, no original accuracy research of the algorithm
Linxweiler 2015	Mucosal swabs for cytological analyses in already diagnosed HNSCC patients (and healthy controls)
Madrigal 2018	Narrative review

Mehrotra 2012	Narrative review
Miller 2017	No seprarate analyses for cell-blocks (cell-block + core biopsy analyzed together)
Miller 2016	Conference abstract
Pai 2009	DTA parameters unobtainable from presented data
Patel 2019	Conference abstract
Pusztaszeri 2015	Narrative review
Qureishi 2017	Narrative review
Rohrback 2017	Case-report of 2 cases
Rollo 2018	Unclear whether this assay is used in the Netherlands
Roy-Chowdhuri 2015	Narrative review
Segura 2018	Conference abstract (via web of science)
Shelton 2017	Tested on biopsies or resected tissue
Shirsat 2018	Conference abstract
Shirsat 2017	Conference abstract
Thanky 2017	Niet te vinden via pubmed, google scholar en web of science. Referentie verwijst naar pagina E97 in vol 39 van Head & Neck, deze (E97 in vol 39) is niet te vinden op de website van het journal: pagina E97 is onderdeel van een case-report van andere auteurs.
Troussier 2018	(Article in French)
Virk 2015	Conference abstract
Wilson 2019	Conference abstract
Wong 2019	Only HPV positive patients were selected.
Xu 2016	conference abstract
Yang 2018	Conference abstract
Yang 2019	Reference test does not seem to be a DNA/mRNA test

Literature search strategy

Histopathology-based diagnostic algorithms

Uitgangsvraag: Welke teststrategie is het meest effectief om de HPV status te bepalen bij patiënten met een orofarynxcarcinoom?	
Database(s): PubMed, Embase	Datum: 10-3-2020
Periode: niet van toepassing	Talen: niet van toepassing
Toelichting: Specifiek gezocht op orofarynx neoplasma. Met uitzondering van het artikel van Smeets worden alle sleutelartikelen gevonden. In het artikel van Smeets wordt gesproken over hoofd hals tumoren en niet specifiek orofarynx tumoren. Er is gezocht naar systematic reviews. Daarnaast is een diagnostisch filter gebruikt. Met vriendelijke groet, Ingeborg van Dusseldorp	

	Inclusief dubbele referenties	Ontdubbeld
SR	47	36
Diagnostisch	573	455
Totaal		**491**

Zoekverantwoording

PubMed

Search	Query	Items found
#62	Search #54 AND #61	275
#61	Search "Sensitivity and Specificity"(MeSH) OR specificit*(tw) OR screening(tw) OR accura*(tw) OR reference value*(tw) OR false positive(tw) OR false negative(tw) OR predictive value*(tw) OR roc(tw) OR likelyhood*(tw) OR likelihood*(tw)	2683505
#57	Search #54 AND #55	23
#55	Search ("Meta-Analysis as Topic"(Mesh) OR "Meta-Analysis"(Publication Type) OR metaanaly*(tiab) OR metanaly*(tiab) OR meta-analy*(tiab) OR meta synthes*(tiab) OR metasynthes*(tiab) OR meta ethnograph*(tiab) OR metaethnograph*(tiab) OR meta summar*(tiab) OR metasummar*(tiab) OR meta-aggregation(tiab) OR metareview(tiab) OR meta-review(tiab) OR overview of reviews(tiab) OR ((systematic*(ti) OR scoping(ti) OR umbrella(ti) OR meta-narrative(ti) OR metanarrative(ti) OR evidence based(ti)) AND (review*(ti) OR overview*(ti))) OR ((evidence(ti) OR narrative(ti) OR metanarrative(ti) OR qualitative(ti)) AND synthesis(ti)) OR systematic review(pt) OR prisma(tiab) OR preferred reporting items(tiab) OR quadas*(tiab) OR systematic review*(tiab) OR systematic literature(tiab) OR structured literature search(tiab) OR systematic overview*(tiab) OR scoping review*(tiab) OR umbrella review*(tiab) OR mapping review*(tiab) OR systematic mapping(tiab) OR evidence synthes*(tiab) OR narrative synthesis(tiab) OR metanarrative synthesis(tiab) OR research synthesis(tiab) OR qualitative synthesis(tiab) OR realist synthesis(tiab) OR realist review(tiab) OR realist evaluation(tiab) OR systematic qualitative review(tiab) OR mixed studies review(tiab) OR mixed methods synthesis(tiab) OR mixed research synthesis(tiab) OR quantitative literature review(tiab) OR systematic evidence review(tiab) OR evidence-based review(tiab) OR comprehensive literature search(tiab) OR integrated review*(tiab) OR integrated literature review(tiab) OR integrative review*(tiab) OR integrative literature review*(tiab) OR structured literature review*(tiab) OR systematic search and review(tiab) OR meta-narrative review*(tiab) OR metanarrative review(tiab) OR systematic narrative review(tiab) OR systemic review(tiab) OR systematized review(tiab) OR systematic research synthesis(tiab) OR bibliographic*(tiab) OR hand-search*(tiab) OR handsearch*(tiab) OR manual search*(tiab) OR searched manually(tiab) OR manually searched(tiab) OR journal database*(tiab) OR review authors independently(tiab) OR reviewers independently(tiab) OR independent reviewers(tiab) OR independent review authors(tiab) OR electronic database search*(tiab) OR (study selection(tiab) AND data extraction(tiab)) OR (selection criteria(tiab) AND data collection(tiab)) OR (selection criteria(tiab) AND data analysis(tiab)) OR (evidence acquisition(tiab) AND evidence synthesis(tiab)) OR (pubmed(tiab) AND embase(tiab)) OR (medline(tiab) AND embase(tiab)) OR (pubmed(tiab) AND cochrane(tiab)) OR (medline(tiab) AND cochrane(tiab)) OR (embase(tiab) AND cochrane(tiab)) OR (pubmed(tiab) AND psycinfo(tiab)) OR (medline(tiab) AND psycinfo(tiab)) OR (embase(tiab) AND psycinfo(tiab)) OR (cochrane(tiab) AND psycinfo(tiab)) OR (pubmed(tiab) AND web of science(tiab)) OR (medline(tiab) AND web of science(tiab)) OR (embase(tiab) AND web of science(tiab)) OR (psycinfo(tiab) AND web of science(tiab)) OR (cochrane(tiab) AND web of science(tiab)) OR ((literature(ti) OR qualitative(ti) OR quantitative(ti) OR integrated(ti) OR integrative(tiab) OR rapid(ti) OR short(ti) OR critical*(ti) OR mixed stud*(ti) OR mixed method*(ti) OR focused(ti) OR focussed(ti) OR structured(ti) OR comparative(ti) OR comparitive(ti) OR evidence(ti) OR comprehensive(ti) OR realist(ti)) AND (review*(ti) OR overview*(ti)) AND (literature search(tiab) OR structured search(tiab) OR electronic search(tiab) OR search strategy(tiab) OR gray literature(tiab) OR grey literature(tiab) OR Review criteria(tiab) OR eligibility criteria(tiab) OR inclusion criteria(tiab) OR exclusion criteria(tiab) OR predetermined criteria(tiab) OR included studies(tiab) OR identified studies(tiab) OR (systematic search(tiab) AND literature(tiab)) OR strength of evidence(tiab) OR citation*(tiab) OR references(tiab) OR database search*(tiab) OR electronic database*(tiab) OR data base search*(tiab) OR electronic data-base*(tiab) OR search criteria(tiab) OR study selection(tiab) OR data extraction(tiab) OR methodological quality(tiab) OR methodological characteristics(tiab) OR methodologic quality(tiab) OR methodologic characteristics(tiab))) OR ((literature review(tiab) OR literature search*(tiab)) AND (structured search(tiab) OR electronic search(tiab) OR Search strategy(tiab) OR gray literature(tiab) OR grey literature(tiab) OR review criteria(tiab) OR eligibility criteria(tiab) OR inclusion criteria(tiab) OR exclusion	335882

	criteria(tiab) OR predetermined criteria(tiab) OR included studies(tiab) OR identified studies(tiab) OR (systematic search(tiab) AND literature(tiab)) OR strength of evidence(tiab) OR citation*(tiab) OR references(tiab) OR database search*(tiab) OR electronic database*(tiab) OR data base search*(tiab) OR electronic data-base*(tiab) OR search criteria(tiab) OR study selection(tiab) OR data extraction(tiab) OR methodological quality(tiab) OR methodological characteristics(tiab) OR methodologic quality(tiab) OR methodologic characteristics(tiab)))) NOT ("Comment" (Publication Type) OR "Letter" (Publication Type)) NOT (“Animals”(Mesh) NOT “Humans”(Mesh))	
#54	Search #44 AND #45 AND #53	1030
#53	Search "Algorithms"(Mesh) OR "Diagnostic Techniques and Procedures"(Mesh) OR algorhythm(tiab) OR algorism(tiab) OR algorithm(tiab) OR algorithms(tiab) OR diagnostic test/exp OR test(tiab) OR tests(tiab)	8648716
#44	Search "Oropharyngeal Neoplasms"(Mesh) OR oropharynx tumor*(tiab) OR oropharynx tumour*(tiab) OR (squamous cell(tiab) AND oropharyn*(tiab)) OR oropharynx cancer(tiab) OR oropharyngeal cancer(tiab) OR oropharynx carcinoma*(tiab) OR oropharyngeal carcinoma*(tiab) OR oropharynx neoplasm*(tiab) OR oropharyngeal neoplasm*(tiab)	11717
#45	Search "Papillomaviridae"(Mesh) OR human papilloma virus(tiab) OR wart virus(tiab) OR condyloma virus(tiab) OR hpv(tiab) OR human papillomavirus(tiab) OR verruca virus(tiab) OR viral verruca(tiab) OR virus verruca(tiab) OR virus wart(tiab)	55608

Embase

No.	Query	Results
#20	#19 NOT #18	**298**
#19	#4 AND #5	**307**
#18	#4 AND #17	**24**
#17	('meta analysis'/de OR cochrane:ab OR embase:ab OR psycinfo:ab OR cinahl:ab OR medline:ab OR ((systematic NEAR/1 (review OR overview)):ab,ti) OR ((meta NEAR/1 analy*):ab,ti) OR metaanalys*:ab,ti OR 'data extraction':ab OR cochrane:jt OR 'systematic review'/de) NOT (('animal experiment'/exp OR 'animal model'/exp OR 'nonhuman'/exp) NOT 'human'/exp)	**481200**
#16	#14 NOT #15	**1**
#15	#6 AND #14	**8**
#14	#7 OR #8 OR #9 OR #10 OR #11 OR #12	**9**
#13	defining AND better AND algorithm AND for AND the AND accurate AND identification AND of AND hpv AND status AND among AND oropharyngeal AND 'squamous cell' AND carcinoma AND 2015	**0**
#12	diagnosis AND of AND hpv AND driven AND head AND neck AND cancer AND comparing AND p16 AND based AND algorithms AND with AND the AND rnascope AND 'hpv test' AND mirghani	**2**
#11	double AND positivity AND is AND the AND biomarker AND with AND strongest AND diagnostic AND accuracy AND prognostic AND value AND for AND human AND papillomavirus AND related AND oropharyngeal AND cancer AND patients AND mena	**1**
#10	developing AND a AND new AND diagnostic AND algorithm AND for AND human AND papillomavirus AND associated AND oropharyngeal AND carcinoma AND cohen AND 2017	**1**

#9	validation AND of AND a AND novel AND diagnostic AND standard AND in AND 'hpv positive' AND oropharyngeal AND squamous AND cell AND carcinoma AND schache AND 2013 AND british AND journal AND cancer	**1**
#8	detection AND human AND papillomavirus AND in AND clinical AND samples AND evolving AND methods AND strategies AND for AND the AND accurate AND determination AND hpv AND status AND of AND head AND neck AND carcinomas AND westra AND 2014	**3**
#7	a AND novel AND algorithm AND for AND reliable AND detection AND of AND human AND papillomavirus AND in AND paraffin AND embedded AND head AND neck AND cancer AND specimen AND smeets	**1**
#6	#4 AND #5	**412**
#5	'sensitivity and specificity'/exp OR 'screening'/exp OR 'reference value'/exp OR 'diagnostic error'/exp OR 'predictive value'/exp OR 'receiver operating characteristic'/exp OR specificit*:ab,ti OR screening:ab,ti OR accura*:ab,ti OR 'reference value*':ab,ti OR 'false positive':ab,ti OR 'false negative':ab,ti OR 'predictive value*':ab,ti OR roc:ab,ti OR likelyhood*:ab,ti OR likelihood*:ab,ti	**3030582**
#4	#1 AND #2 AND #3	**1378**
#3	'algorithm'/exp OR 'algorhythm*':ti,ab,kw OR 'algorism*':ti,ab,kw OR 'algorithm*':ti,ab,kw OR 'diagnostic test'/exp OR 'test battery'/exp OR test:ti,ab,kw OR tests:ti,ab,kw OR 'assay combination':ti,ab,kw	**3687381**
#2	'wart virus'/exp OR 'human papilloma virus':ti,ab OR 'wart virus':ti,ab OR 'condyloma virus':ti,ab OR 'hpv':ti,ab OR 'human papillomavirus':ti,ab OR 'verruca virus':ti,ab OR 'viral verruca':ti,ab OR 'virus verruca':ti,ab OR 'virus wart':ti,ab OR 'verruca, viral':ti,ab OR 'papillomaviridae'/exp	**77439**
#1	'oropharynx cancer'/exp OR ((oropharyn* NEAR/3 (tumor* OR 'squamous cell' OR tumour* OR cancer* OR carcinoma* OR neoplasm*)):ti,ab,kw)	**14838**

Cytology-based tests for positive neck nodes and CUP

Algemene informatie

Richtlijn: Hoofd- halstumoren	
Uitgangsvraag: Hoe dient de HPV-status bepaald te worden?	
Database(s): PubMed, Embase	Datum: 27-2-2020
Periode: niet van toepassing	Talen: niet van toepasing
Literatuurspecialist: Ingeborg van Dusseldorp	
Toelichting en opmerkingen: Voor deze vraag is op een specifieke manier gezocht conform de opgave in het zoekformulier. Dat betekent concreet dat er een combinatie is gemaakt van: HHT tumoren EN cytologie, fine needle aspiration EN p16 staining, immunohistochemy et cetera zie zoekformulier EN HPV Vervolgens zijn de systematische reviews geselecteerd en is een combinatie gemaakt met een sensitief diagnostisch filter. Bij het specifieke diagnostische filter werd het sleutelartikel van Rollo niet gevonden, vandaar de keuze voor het sensitieve filter, waarmee alle sleutelartikelen werden gevonden.	

Zoekopbrengst

	Embase	PubMed	Ontdubbeld
SRs	6	8	13
RCTs			
Observationele studies			
Diagnostische studies	285	201	435

Totaal			448

Zoekverantwoording

PubMed

Search	Query	Results
#19	Search: #12 AND #17	201
#15	Search: #10 AND #13	8
#17	Search: "Immunohistochemistry"(Mesh) OR “22c3 pharmdx”(tiab) OR “28-8 pharmdx”(tiab) OR “er-pr pharmdx”(tiab) OR “envision duoflex”(tiab) OR “envision flex”(tiab) OR “pd-l1 22c3 pharmdx”(tiab) OR “pd-l1 ihc 22c3 pharmdx”(tiab) OR “pd-l1 ihc 28-8 pharmdx”(tiab) OR “ventana pd-l1”(tiab) OR “confirm id kit”(tiab) OR “immunohistochemical test kit”(tiab) OR pharmdx(tiab) OR antigen staining(tiab) OR immunohistochemistr*(tiab) OR immunostaining(tiab) OR "in situ hybridization"(MeSH Terms) OR “hybridization in situ”(tiab) OR “in situ hybridisation”(tiab) OR “hybridisation in situ”(tiab) OR "Polymerase Chain Reaction"(Mesh) OR “polymerase chain reaction*”(tiab) OR p16(tiab) OR "P16 protein, human" (Supplementary Concept)	1,274,482
#13	Search: ("Meta-Analysis as Topic"(Mesh) OR “Meta-Analysis”(Publication Type) OR metaanaly*(tiab) OR metanaly*(tiab) OR meta-analy*(tiab) OR meta synthes*(tiab) OR metasynthes*(tiab) OR meta ethnograph*(tiab) OR metaethnograph*(tiab) OR meta summar*(tiab) OR metasummar*(tiab) OR meta-aggregation(tiab) OR metareview(tiab) OR meta-review(tiab) OR overview of reviews(tiab) OR ((systematic*(ti) OR scoping(ti) OR umbrella(ti) OR meta-narrative(ti) OR metanarrative(ti) OR evidence based(ti)) AND (review*(ti) OR overview*(ti))) OR ((evidence(ti) OR narrative(ti) OR metanarrative(ti) OR qualitative(ti)) AND synthesis(ti)) OR systematic review(pt) OR prisma(tiab) OR preferred reporting items(tiab) OR quadas*(tiab) OR systematic review*(tiab) OR systematic literature(tiab) OR structured literature search(tiab) OR systematic overview*(tiab) OR scoping review*(tiab) OR umbrella review*(tiab) OR mapping review*(tiab) OR systematic mapping(tiab) OR evidence synthes*(tiab) OR narrative synthesis(tiab) OR metanarrative synthesis(tiab) OR research synthesis(tiab) OR qualitative synthesis(tiab) OR realist synthesis(tiab) OR realist review(tiab) OR realist evaluation(tiab) OR systematic qualitative review(tiab) OR mixed studies review(tiab) OR mixed methods synthesis(tiab) OR mixed research synthesis(tiab) OR quantitative literature review(tiab) OR systematic evidence review(tiab) OR evidence-based review(tiab) OR comprehensive literature search(tiab) OR integrated review*(tiab) OR integrated literature review(tiab) OR integrative review*(tiab) OR integrative literature review*(tiab) OR structured literature review*(tiab) OR systematic search and review(tiab) OR meta-narrative review*(tiab) OR metanarrative review(tiab) OR systematic narrative review(tiab) OR systemic review(tiab) OR systematized review(tiab) OR systematic research synthesis(tiab) OR bibliographic*(tiab) OR hand-search*(tiab) OR handsearch*(tiab) OR manual search*(tiab) OR searched manually(tiab) OR manually searched(tiab) OR journal database*(tiab) OR review authors independently(tiab) OR reviewers independently(tiab) OR independent reviewers(tiab) OR independent review authors(tiab) OR electronic database search*(tiab) OR (study selection(tiab) AND data extraction(tiab)) OR (selection criteria(tiab) AND data collection(tiab)) OR (selection criteria(tiab) AND data analysis(tiab)) OR (evidence acquisition(tiab) AND evidence synthesis(tiab)) OR (pubmed(tiab) AND embase(tiab)) OR (medline(tiab) AND embase(tiab)) OR	334,389

	(pubmed(tiab) AND cochrane(tiab)) OR (medline(tiab) AND cochrane(tiab)) OR (embase(tiab) AND cochrane(tiab)) OR (pubmed(tiab) AND psycinfo(tiab)) OR (medline(tiab) AND psycinfo(tiab)) OR (embase(tiab) AND psycinfo(tiab)) OR (cochrane(tiab) AND psycinfo(tiab)) OR (pubmed(tiab) AND web of science(tiab)) OR (medline(tiab) AND web of science(tiab)) OR (embase(tiab) AND web of science(tiab)) OR (psycinfo(tiab) AND web of science(tiab)) OR (cochrane(tiab) AND web of science(tiab)) OR ((literature(ti) OR qualitative(ti) OR quantitative(ti) OR integrated(ti) OR integrative(tiab) OR rapid(ti) OR short(ti) OR critical*(ti) OR mixed stud*(ti) OR mixed method*(ti) OR focused(ti) OR focussed(ti) OR structured(ti) OR comparative(ti) OR comparitive(ti) OR evidence(ti) OR comprehensive(ti) OR realist(ti)) AND (review*(ti) OR overview*(ti)) AND (literature search(tiab) OR structured search(tiab) OR electronic search(tiab) OR search strategy(tiab) OR gray literature(tiab) OR grey literature(tiab) OR Review criteria(tiab) OR eligibility criteria(tiab) OR inclusion criteria(tiab) OR exclusion criteria(tiab) OR predetermined criteria(tiab) OR included studies(tiab) OR identified studies(tiab) OR (systematic search(tiab) AND literature(tiab)) OR strength of evidence(tiab) OR citation*(tiab) OR references(tiab) OR database search*(tiab) OR electronic database*(tiab) OR data base search*(tiab) OR electronic data-base*(tiab) OR search criteria(tiab) OR study selection(tiab) OR data extraction(tiab) OR methodological quality(tiab) OR methodological characteristics(tiab) OR methodologic quality(tiab) OR methodologic characteristics(tiab))) OR ((literature review(tiab) OR literature search*(tiab)) AND (structured search(tiab) OR electronic search(tiab) OR Search strategy(tiab) OR gray literature(tiab) OR grey literature(tiab) OR review criteria(tiab) OR eligibility criteria(tiab) OR inclusion criteria(tiab) OR exclusion criteria(tiab) OR predetermined criteria(tiab) OR included studies(tiab) OR identified studies(tiab) OR (systematic search(tiab) AND literature(tiab)) OR strength of evidence(tiab) OR citation*(tiab) OR references(tiab) OR database search*(tiab) OR electronic database*(tiab) OR data base search*(tiab) OR electronic data-base*(tiab) OR search criteria(tiab) OR study selection(tiab) OR data extraction(tiab) OR methodological quality(tiab) OR methodological characteristics(tiab) OR methodologic quality(tiab) OR methodologic characteristics(tiab)))) NOT ("Comment" (Publication Type) OR "Letter" (Publication Type)) NOT (“Animals”(Mesh) NOT “Humans”(Mesh))	
#12	Search: #10 AND #11	239
#11	Search: "Sensitivity and Specificity"(MeSH) OR "Diagnostic Errors"(MeSH) OR sensitive(tw) OR sensitivity(tw) OR specificity(tw) OR accurate(tw) OR accuracy(tw) OR "golden standard" OR "gold standard" OR (reference(tw) AND (test(tw) OR standard(tw))) OR "index test" OR validity(tw) OR validation(tw) OR validate*(tw) OR valid(ti) OR validation studies(pt) OR verif*(ti) OR evaluation studies(pt) OR evaluat*(ti) OR (false(tw) AND (positive(tw) OR negative(tw))) OR pretest(tw) OR pre-test(tw) OR posttest(tw) OR post-test(tw) OR predictive value OR predict*(ti) OR roc(tw) OR likelyhood(tw) OR likelihood(tw) OR value*(ti) OR reference values(mesh) OR cutoff(tw) OR cut-off(tw) OR quality control(mesh) OR "reproducibility of results"(mesh) OR repeatability(tw) OR reproducibility(tw) OR efficacy(tw) OR reliability(tw) OR comparative study(pt) OR odds(tw) OR error*(tw) OR suitability(tw) OR utility(tw)	7,058,350
#10	Search: #6 AND #7 AND #9	535
#9	Search: "cytology" (Subheading) OR "Cytological Techniques"(Mesh) OR "Cell Biology"(Mesh) OR “automated cytological technique*”(tiab) OR cell biology(tiab) OR cytolog*(tiab) OR cytotechnolog*(tiab) OR cytotest*(tiab) OR "Biopsy, Fine-	2,591,742

	Needle"(Mesh) OR fine core needle biop*(tiab) OR fine needle aspiration*(tiab) OR fine needle biop*(tiab) OR needle aspiration*(tiab) OR fnab(tiab) OR fnac(tiab)	
#7	Search: Head and Neck Neoplasms(Mesh:NoExp) OR Squamous Cell Carcinoma of Head and Neck(Mesh) OR neck mass*(tiab) OR neck tumor*(tiab) OR neck tumour*(tiab) OR neck cancer*(tiab)	67,320
#6	Search: "Papillomaviridae"(Mesh) OR human papilloma virus(tiab) OR wart virus(tiab) OR condyloma virus(tiab) OR hpv(tiab) OR human papillomavirus(tiab) OR verruca virus(tiab) OR viral verruca(tiab) OR virus verruca(tiab) OR virus wart(tiab)	55,513

Embase

Embase Session Results (27 Feb 2020)

No.	Query	Results
#16	#12 AND #14	**285**
#15	#1 AND #14	**6**
#14	#7 AND #8 AND #10 AND #13	**302**
#13	'immunohistochemical test kit'/exp OR '22c3 pharmdx':ti,ab,kw OR '28-8 pharmdx':ti,ab,kw OR 'er-pr pharmdx':ti,ab,kw OR 'envision duoflex':ti,ab,kw OR 'envision flex':ti,ab,kw OR 'pd-l1 22c3 pharmdx':ti,ab,kw OR 'pd-l1 ihc 22c3 pharmdx':ti,ab,kw OR 'pd-l1 ihc 28-8 pharmdx':ti,ab,kw OR 'ventana pd-l1':ti,ab,kw OR 'confirm id kit':ti,ab,kw OR 'immunohistochemical test kit':ti,ab,kw OR 'immunohistochemistry test kit':ti,ab,kw OR 'pharmdx':ti,ab,kw OR 'immunohistochemistry'/exp OR 'antigen staining':ti,ab,kw OR 'immunohistochemistr*':ti,ab,kw OR 'immunostaining':ti,ab,kw OR 'in situ hybridization'/exp OR 'hybridization in situ':ti,ab,kw OR 'in situ hybridisation':ti,ab,kw OR 'hybridisation in situ':ti,ab,kw OR 'polymerase chain reaction'/exp OR 'polymerase chain reaction':ti,ab,kw OR p16:ti,ab,kw OR 'protein p16'/exp	**1557687**
#12	sensitiv* OR detect* OR accura* OR specific* OR reliab* OR positive OR negative OR diagnos*	**14295362**
#11	#9 AND #10	**500**
#10	'cytology'/exp OR 'automated cytological technique*':ti,ab,kw OR 'cell biology':ti,ab,kw OR 'cytolog*':ti,ab,kw OR 'cytotechnolog*':ti,ab,kw OR 'cytotest*':ti,ab,kw OR 'fine needle aspiration biopsy'/exp OR 'fine core needle biop*':ti,ab,kw OR 'fine needle aspiration*':ti,ab,kw OR 'fine needle biop*':ti,ab,kw OR 'needle aspiration*':ti,ab,kw OR fnab:ti,ab,kw OR fnac:ti,ab,kw	**980450**
#9	#7 AND #8	**6886**
#8	'neck tumor'/exp OR 'neck mass*':ti,ab,kw OR 'neck tumor*':ti,ab,kw OR 'neck tumour*':ti,ab,kw OR 'neck cancer*':ti,ab,kw OR 'head and neck carcinoma'/exp OR 'neck carcinoma*':ti,ab,kw	**114000**
#7	'wart virus'/exp OR 'human papilloma virus':ti,ab OR 'wart virus':ti,ab OR 'condyloma virus':ti,ab OR 'hpv':ti,ab OR 'human papillomavirus':ti,ab OR 'verruca virus':ti,ab OR 'viral verruca':ti,ab OR 'virus verruca':ti,ab OR 'virus wart':ti,ab OR 'verruca, viral':ti,ab OR 'papillomaviridae'/exp	**77320**
#6	#2 OR #3 OR #4 OR #5	**4**
#5	27816020	**1**
#4	29873841	**1**
#3	31246356	**1**
#2	role AND cytology AND in AND the AND diagnosis AND management AND of AND 'hpv associated' AND head AND neck AND carcinoma	**1**

#1	('meta analysis'/de OR cochrane:ab OR embase:ab OR psycinfo:ab OR cinahl:ab OR medline:ab OR ((systematic NEAR/1 (review OR overview)):ab,ti) OR ((meta NEAR/1 analy*):ab,ti) OR metaanalys*:ab,ti OR 'data extraction':ab OR cochrane:jt OR 'systematic review'/de) NOT (('animal experiment'/exp OR 'animal model'/exp OR 'nonhuman'/exp) NOT 'human'/exp)	**479733**

Module 10.1 Aanvraag en verslag pathologie-onderzoek

Uitgangsvraag

Wat moet er minimaal in de aanvraag voor en in het verslag van pathologie-onderzoek staan?

Inleiding

Het pathologieverslag dient alle informatie te bevatten die noodzakelijk is voor het bepalen van de therapie door de behandelaar, zowel in het geval van een biopt, als in het geval van een resectie. Daarom is het van belang dat deze informatie door de patholoog op een gestructureerde wijze wordt verwerkt in het pathologieverslag.

Samenvatting literatuur

Er is geen systematische literatuursearch uitgevoerd om de uitgangsvraag te kunnen beantwoorden. Tijdens de voorbereiding heeft de werkgroep afgesproken om aan te sluiten bij het relevante PALGA-protocol (Stichting PALGA, 2019). Idealiter zou de werkgroep beschikken over informatie ('evidence') over de (kosten)effectiviteit van het gebruik van het huidige PALGA protocol in de Nederlandse praktijk. Dergelijke effectiviteitsstudies zijn volgens de werkgroep niet gedaan en het is niet aannemelijk dat de werkgroep dergelijke studies over het hoofd zou hebben gezien. Vanuit het oogpunt van doelmatig gebruik van beschikbare tijd en middelen voor de richtlijnontwikkeling heeft de werkgroep besloten om geen systematische literatuursearch uit te (laten) voeren voor de onderbouwing van de aanbevelingen van deze module.

Overwegingen

Algemeen

Het beleid bij de behandeling van hoofd-halstumoren wordt mede bepaald door het type tumor. De meeste in het hoofd-halsgebied voorkomende tumoren betreffen mucosale plaveiselcelcarcinomen. Daarnaast komen andere typen carcinomen voor in de speekselklieren en neus(bijholte). Zowel speekselkliercarcinomen als sino-nasale carcinomen zijn relatief zeldzaam. De meeste literatuur over prognostische tumorkenmerken geldt voor de mucosale plaveiselcelcarcinomen.

Plaveiselcelcarcinoom van het hoofd-halsgebied

Na een biopt of resectie voor een hoofd-halscarcinoom wordt het verdere therapeutische beleid in eerste instantie mede bepaald op basis van de WHO-classificatie en stadiering van de tumor (Barnes, 2005). Derhalve is het cruciaal dat in het PA verslag van de resectie alle componenten van de TN classificatie volledig worden benoemd. Daarnaast zijn er parameters die prognostische waarde hebben, onafhankelijk van de classificatie, die mede bepalend zijn voor het postoperatief beleid. Dit betreft karakteristieken van de tumor en metastasen die uitsluitend middels microscopisch onderzoek kunnen worden vastgesteld (Barnes, 2005; Odell, 1994; Woolgar, 1999, 2005, 2006 en 2009; Pentenero, 2005; Bradley, 2007; Weijers, 2009; Jones, 2009; Huang, 2009; Brandwein-Gensler 2010; Coca-Pelaz, 2012; Jerjes, 2012; Brown, 2012). Om tot een beslissing over het postoperatief beleid bij een patiënt te komen, moeten al deze ingrediënten in het pathologieverslag eenduidig benoemd staan. In het verslag staan zowel alle ingrediënten die nodig zijn voor het vaststellen van de pTN classificatie als de prognostische parameters die het postoperatief beleid mede bepalen. Bij gebruik van de PALGA protocolmodule is de volledigheid en uniformiteit van het verslag gewaarborgd.

Sommige aanvragen van bepalingen zijn vooralsnog niet standaard, maar kunnen wellicht op indicatie aangevraagd worden in specifieke klinische settings. Zo is te denken aan een immunohistochemische bepaling van PD-L1 eiwitten als indicatie voor pembrolizumab, welke aan PD-L1 scores is gekoppeld en wordt uitgedrukt in het zogenoemde 'combined positive score' (CPS).

HPV gerelateerde hoofd-halstumoren
Een deel van de carcinomen in de orofarynx zijn humaan papillomavirus (HPV) gerelateerd (Hobbs, 2006; Smeets, 2007; Lechner, 2013; Holzinger, 2013; Ragin and Taioli, 2007; Westra, 2012; Robinson, 2012). Dikwijls wordt een orofarynxcarcinoom behandeld met (chemo)radiatie. Derhalve is een biopt het enige materiaal dat bij deze tumoren wordt verkregen. Omdat de huidige TNM classificatie (8^{e} ed) verschillend is voor p16positieve en -negatieve oropharynxcarcinomen, moet in geval van een biopt uit de orofarynx (tonsil, tongbasis, orofarynx nos) standaard een HPV-bepaling verricht worden. P16 immunohistochemie is een goede surrogaatmarker voor "high-risk"HPV typen. Zie voor de overige aanbevelingen met betrekking tot HPV bepaling de richtlijnmodule HPV-statusbepaling in module 'Diagnostiek Orofarynxcarcinoom'.

De onderstaande minimale histologische dataset voor plaveiselcelcarcinoom is een bewerking van de richtlijnen van de Royal College of Pathology, in gebruik binnen de British Society of Head and Neck Oncology (https://www.rcpath.org/profession/guidelines/cancer-datasets-and-tissue-pathways.html)

PA data resectiepreparaten
1. histologisch (sub)type en differentiatiegraad: subtypen van plaveiselcelcarcinomen (bijvoorbeeld basaloid, verruceus, sarcomatoid, papillair (Barnes, 2005) moeten worden vermeld. Differentiatiegraad wordt vermeld volgens WHO (Barnes, 2005; Weijers, 2009). Indien binnen een tumor variatie is in differentiatiegraad, wordt de hoogste graad (slechtste differentiatie) vermeld. Bij aanwezigheid van uitsluitend een premaligne laesie, wordt de hoogste mate van dysplasie vermeldt volgens WHO (Barnes, 2005);
2. groeipatroon: het groeipatroon in het tumorfront heeft prognostische waarde (Odell, 1994; Woolgar, 1999, 2006; Coca-Pelaz, 2012; Jerjes, 2012; Brown, 2012). Hierbij wordt onderscheid gemaakt in expansieve versus sprieterige groei, waarbij ook een (kleine) component van de tumor met een sprieterig groeipatroon moet worden vermeld. In de PPMI worden 5 typen groeipatronen onderscheiden die daarnaast (optioneel in de module) apart vermeld kunnen worden (Heeremam 2015);
3. tumor diameter: de macroscopisch gemeten tumordiameter wordt vermeld tenzij de microscopische uitbreiding groter is dan macroscopisch werd gemeten. De tumordiameter bepaalt het T stadium;
4. invasiediepte: de invasiediepte wordt gemeten vanaf het mucosale oppervlak. In geval van ge-ulcereerde tumoren wordt het gereconstrueerde mucosale oppervlak als referentiepunt genomen. Invasiediepte van orale carcinomen is gerelateerd aan de kans op lymfkliermetastasen (Pentenero, 2005; Huang, 2009);
5. afstand ten opzichte van de chirurgische resectievlakken: vanuit chirurgisch opzicht is > 5 mm radicaal, 1 tot 5 mm krap radicaal en < 1 mm irradicaal (Woolgar, 2005, 2006; Bradley, 2007; Coca-Pelaz, 2012; Jerjes, 2012; Brown, 2012);
6. vasculaire invasie: hierbij is het niet van belang of het bloedvat- of lymfvatinvasie betreft (Jones, 2009; Brandwein-Gensler, 2010; Jerjes, 2012). Dit moet wel onderscheiden worden van retractie-artefacten rond de tumorvelden;

7. perineurale groei: bij perineurale groei aan het tumorfront is er een grotere kans op lokaal recidief of lymfkliermetastasen (Coca-Pelaz, 2012; Jerjes, 2012);
8. dysplasie in de mucosale resectieranden: gradering volgens WHO (Barnes, 2005);
9. bot/kraakbeeninvasie: hierbij dient een onderscheid te worden gemaakt tussen boterosie waarbij uitsluitend usurering van de cortex aanwezig is en botinvasie waarbij de tumor de cortex volledig doorbreekt. Botinvasie bepaalt de tumorstadiering. In het geval van larynxtumoren bepaalt de invasie van het larynxskelet (kraakbeen) het tumorstadium. Invasie van epiglottis kraakbeen en cart. arythenoidea gelden niet als larynxskelet.

Speekselkliercarcinomen
Speekselkliercarcinomen worden geclassificeerd volgens de WHO (Barnes, 2005). Hierin worden 24 typen maligne tumoren onderscheiden. Voor de prognose is het tumorstadium belangrijker dan het histologisch subtype (Regis De Brito Santos, 2001; Terhaard, 2001; Van der Schroeff, 2010). Er is, qua histologische typering een indeling gemaakt tussen speekselkliercarcinomen van lage en hoge maligniteitsgraad (Speight, 2009) en dit geeft ook een indicatie voor de prognose. Bij sommige speekselkliertumoren wordt gegradeerd, bijvoorbeeld het mucoepidermoïd carcinoom in hoog-, midden- en laaggradig, of het adenoïd cysteus carcinoom qua groeiwijze in cribriform, tubulair, of solide (zie bijlage 1: gradering mucoepidermoid carcinoom) (Barnes,, 2005; Van der Schroeff, 2010; Speight, 2009).

Voor de prognose is radicaliteit van de resectie een criterium (Terhaard, 2001). Postoperatieve radiotherapie is onder andere geïndiceerd bij niet-radicale resecties en uitgebreide perineurale groei in “named nerves” (Barrett, 2009).

Neus, neusbijholte
Sinonasale carcinomen worden geclassificeerd volgens de WHO (Barnes, 2005). Er worden diverse, epitheliale en non-epitheliale maligne tumoren onderscheiden (Barnes, 2005; Ejaz, 2005; Mendenhall, 2006; Renner, 2007; Stelow, 2008; Khademi, 2009; LLorente, 2009; Thompson, 2009; Stelow, 2010; Ansa, 2013; Haerle, 2013; Van Gompel, 2012; Slootweg, 2013). Voor de prognose is het histologisch type en de tumoruitbreiding van belang. In het geval van een adenocarcinoom is het onderscheid tussen intestinaal en non-intestinaal subtype en de differentiatiegraad van prognostische waarde (Llorente, 2009; Thompson, 2009). In het geval van een esthesioneuroblastoom wordt de prognose mede bepaald door Hyam’s gradering (zie bijlage 2: gradering esthesioneuroblastoom) (Thompson, 2009; Van Gompel, 2012). Postoperatieve radiotherapie is geïndiceerd bij niet-radicale resecties en uitgebreide perineurale groei in “named nerves”.

Lymfklierdissecties
a. Halsklierdissectie
Omdat de identificatie van de klierniveaus na uitname van het operatiepreparaat lastig of zelfs onmogelijk kan zijn, worden de klierniveaus idealiter aan het aangeleverde preparaat gemarkeerd door de chirurg (Ferlito, 2002).

Halsklierdissecties worden per level benoemd in het verslag. Van elk level wordt vermeld:
- aantal lymfklieren;
- aantal lymfklieren met metastasen;
- diameter van de grootste metastase (NB: dit is niet de diameter van de lymfklier); bij de diameter van de metastase wordt onderscheid gemaakt in “Isolated tumor cells”

(ITC) bij een metastase ≤ 0,2 mm, micrometastase (> 0,2 mm, < 2 mm) of macrometastase ≥ 2 mm);
- aan- of afwezigheid van extranodale groei.

Extranodale groei is een prognostisch ongunstig kenmerk. Hierbij is het niet van belang of het microscopische of macroscopische kapseldoorbraak betreft. In geval van twijfel wordt de metastase geclassificeerd als hebbende kapseldoorgroei (Snow, 1982; Ferlito., 2002; Puri, 2003; Oosterkamp, 2006; Wan, 2012; Woolgar, 2013). De aanwezigheid van extranodale groei heeft consequenties voor de keuze van de adjuvante behandeling.

b. Schildwachtklierprocedure

Een schilwachtklierprocedure kan worden verricht bij een klinisch negatieve hals (cN0) bij een patiënt met een klein mondholtecarcinoom dat intra-oraal kan worden verwijderd (Gurney, 2012; Sloan, 2009; Alkureishi, 2010; Trivedi, 2010; Broglie, 2011, 2013 Schilling, 2019). De clinicus vermeldt duidelijk op de aanvraag welke klier(en) volgens de schildwachtklierprocedure moeten worden bewerkt.

De aangeleverde schildwachtklier(en) wordt /worden volledig voor histopathologisch onderzoek ingesloten. Klieren > 0,5 cm worden gelamelleerd in plakjes van circa 0,3 cm. Hiervan worden zes niveaus gesneden met een onderlinge afstand van 150 mu. Van elk niveau wordt een HE coupe vervaardigd en een immunohistochemische kleuring met behulp van een monoclonale antistof gericht tegen pankeratine, zoals AE1/3 (Sloan, 2009; Trivedi, 2010; Broglie, 2013).

Bij de diameter van de metastase wordt onderscheid gemaakt in “Isolated tumor cells” (ITC) bij een metastase ≤0,2 mm, micrometastase (> 0,2 mm, < 2 mm) of macrometastase ≥ 2 mm). Zogenaamde “ghostcells” of “mummified cells” (keratine positieve elementen zonder kern) worden niet beschouwd als tumormetastase (Woolgar, 2013).

Aanbevelingen

Vermeld op het aanvraagformulier voor algemeen pathologie-onderzoek het volgende:
De clinicus vermeldt **altijd** de volgende gegevens:
- ingreep;
- tumorlokalisatie;
- zijdigheid
- klinische TN classificatie;
- relevante voorgeschiedenis (eerdere behandeling, relevante comorbiditeit).

De clinicus vermeldt bij halsklierdissectie **aanvullend** de volgende gegevens:
- type halsklierdissectie;
- levels van het halsklierdissectiepreparaat;
- markering van levels op preparaat of bijgevoegde tekening.

De clinicus vermeldt bij een schildwachtklierprocedure **aanvullend** de volgende gegevens:
- welke klier(en) volgens de schildwachtklierprocedure bewerkt moeten worden.

Vermeld in het verslag pathologie-onderzoek onderstaande en **gebruik** hierbij bij voorkeur de PALGA protocol module:
De patholoog vermeldt minimaal de volgende algemene gegevens:
- naam aanvrager;
- datum;
- data patiënt;

- klinische gegevens (letterlijk overgenomen van het aanvraagformulier);
- naam patholoog;
- macroscopie;
- microscopie;
- conclusie.

Specifieke gegevens verslag *mucosaal plaveiselcelcarcinoom*
Het verslag van een *biopt* bevat minimaal de volgende gegevens:
- histologisch type;
- differentiatiegraad optioneel;
- bij orofarynxbiopt: HPV bepaling.

Het verslag van een *resectiepreparaat* bevat minimaal de volgende gegevens:
- aard preparaat, zijdigheid;
- plaats tumor;
- histologisch type en differentiatiegraad tumor;
- groeipatroon;
- maximale diameter;
- invasiediepte;
- kleinste afstand carcinoom tot de resectievlakken in mm, zowel mucosale- als weke delen resectievlakken;
- aan- of afwezigheid (lymf)vaat-ingroei;
- aan- of afwezigheid perineurale groei ter plaatse van het invasieve front;
- aan- of afwezigheid (ernstige) dysplasie in de mucosale resectievlakken;
- indien van toepassing:
- aan- of afwezigheid kraakbeen- of botinvasie;
- aan- of afwezigheid doorgroei in de schildklier.
- aan- of afwezigheid humaan papillomavirus (HPV) bij plaveiselcelcarcinomen van de orofarynx.

Specifieke gegevens verslag *speekselkliercarcinomen*
Het verslag van een *resectiepreparaat* bevat minimaal de volgende gegevens:
- aard preparaat, zijdigheid;
- histologisch tumortype en indien relevant (bijvoorbeeld mucoepidermoïd carcinoom) gradering;
- maximale diameter;
- ingroei in omgevende structuren;
- perineurale groei in “named nerves”;
- kleinste afstand carcinoom tot de resectievlakken in mm.

Specifieke gegevens verslag *neus- en neusbijholtetumoren*
Het verslag van een *resectiepreparaat* bevat minimaal de volgende gegevens:
- aard preparaat, zijdigheid;
- histologisch tumortype en indien relevant (bijvoorbeeld esthesioneuroblastoom) gradering;
- maximale diameter;
- ingroei in omgevende structuren (bot);
- perineurale groei in “named nerves”;
- Kleinste afstand carcinoom tot de resectievlakken in mm, zowel mucosale- als weke delen resectievlakken.

Specifieke gegevens verslag *lymfeklierresectie*
Het verslag van een *lymfeklierresectiepreparaat* van de hals bij een primair hoofd-hals carcinoom of bij een metastase ven onbekende primaire tumor bevat minimaal de volgende gegevens:

- aard preparaat (type halsklierdissectie met vermelding levels);
- per level:
- aantal lymfklieren;
- aantal lymfklieren met metastasen;
- diameter van de grootste metastase;
- aan- of afwezigheid van extranodale groei.

Specifieke gegevens verslag *schildwachtklierprocedure*
Het verslag van een *schildwachtklierprocedure* bevat minimaal de volgende gegevens:

- aantal lymfklieren;
- aan- of afwezigheid van een metastase;
- diameter van de metastase (Isolated Tumor Cells/micro-/macrometastasen);
- aan- of afwezigheid van extranodale groei.

Literatuur

Alkureishi LW, Ross GL, Shoaib T, et al. (2010). Sentinel node biopsy in head and neck squamous cell cancer: 5-year follow-up of a European multicenter trial. Ann SurgOncol, 17(9), 2459-64.

Ansa B, Goodman M, Ward K, et al.(2013). Paranasal sinus squamous cell carcinoma incidence and survival based on Surveillance, Epidemiology, and End Results data, 1973 to 2009. Cancer, 119(14), 2602-10.

Barnes L, Eveson JW, Sidransky D. (2005). IARC Press, Lyon, WHO Classification of tumours. Pathology and Genetics. Head and Neck tumours. Ed.

Barrett AW, Speight PM. (2009). Perineural invasion in adenoid cystic carcinoma of the salivary glands: a valid prognostic indicator? Oral Oncol, 45, 936-40.

Bradley PJ, MacLennan K, Brakenhoff RH, et al. (2007). Status of primary tumour surgical margins in squamous head and neck cancer: prognostic implications. CurrOpinOtolaryngol Head Neck Surg, 15, 74–81.

Brandwein MS, Ivanov K, Wallace I, et al. (2001) Mucoepidermoid carcinoma: a clinicopathologic study of 80 patients with special reference to histologic grading. Am J SurgPathol, 25, 835–845.

Brandwein-Gensler M, Smith RV. (2010). Prognostic indicators in head and neck oncology including the new 7th edition of the AJCC staging system. Head Neck Pathol, 4, 53–61.

Broglie MA, Haerle SK, Huber GF, et al. (2013). Occult metastases detected by sentinel node biopsy in patients with early oraland oropharyngeal squamous cell carcinomas: Impact on survival. Head Neck, 35(5), 660-6.

Broglie MA, Haile SR, Stoeckli SJ. (2011). Long-term experience in sentinel node biopsy for early oral and oropharyngeal squamous cell carcinoma. Ann SurgOncol, 18(10), 2732-8.

Brown JS, Shaw RJ, Bekiroglu F, et al. (2012). Systematic review of the current evidence in the use of postoperative radiotherapy for oral squamous cellcarcinoma. Br J Oral Maxillofac Surg., 50(6), 481-9.

Coca-Pelaz A, Rodrigo JP, Suárez C. (2012). Clinicopathologic analysis and predictive factors for distant metastases inpatients with head and neck squamous cell carcinomas.Head Neck., 34(6), 771-5.

Ejaz A, Wenig BM. (2005). Sinonasal undifferentiated carcinoma: clinical and pathologic features and a discussion on classification, cellular differentiation, and differential diagnosis. AdvAnatPathol, 12, 134-43.
Ferlito A, Rinaldo A, Devaney KO, et al. (2002). Prognostic significance of microscopic and macroscopic extracapsular spread from metastatic tumor in the cervical lymph nodes.Oral Oncol, 38(8), 747-51.
Ferlito A, Robbins KT, Shah JP, et al. (2011). Proposal for a rational classification of neck dissections. Head Neck., 33(3), 445-50.
Gurney BA, Schilling C, Putcha V, et al. (2012). Implications of a positive sentinel node in oral squamous cell carcinoma. Head Neck, 34(11), 1580-5.
Haerle SK, Gullane PJ, Witterick IJ, et al. (2013)., Sinonasal carcinomas: epidemiology, pathology, and management. NeurosurgClin N Am, 24(1), 39-49.
Heerema MG, Melchers LJ, Roodenburg JL, Schuuring E, de Bock GH, & van der Vegt B. (2016). Reproducibility and prognostic value of pattern of invasion scoring in low-stage oral squamous cell carcinoma. Histopathology, 68(3), 388-97.
Hobbs CG, Sterne JA, Bailey M, et al. (2006). Human papillomavirus and head and neck cancer: a systematic review and meta-analysis. ClinOtolaryngol, 31, 259–266.
Huang SH, Hwang D, Lockwood G, et al. (2009). Predictive value of tumor thickness for cervical lymph-node involvement in squamous cell carcinoma of the oral cavity: a meta-analysis of reported studies. Cancer, 115, 1489–1497.
Jerjes W, Upile T, Petrie A, et al. (2012). Clinicopathological parameters, recurrence, locoregional and distant metastasis in 115 T1-T2 oralsquamous cell carcinoma patients. Head Neck Oncol.,2, 9-21.
Jones HB, Sykes A, Bayman N, et al. (2009). The impact of lymphovascular invasion on survival in oral carcinoma.Oral Oncol., 45(1), 10-5.
Khademi B, Moradi A, Hoseini S, et al. (2009). Malignant neoplasms of the sinonasal tract: report of 71 patients and literature review and analysis. Oral Maxillofac Surg., 13, 191-9.
Llorente JL, Perez-Escuredo J, Alvarez-Marcos C, et al. (2009). Genetic and clinical aspects of wood dust related intestinal-type sinonasal adenocarcinoma: a review. Eur Arch Otorhinolaryngol, 266, 1–7.
Mendenhall WM, Mendenhall CM, Riggs CE Jr, et al. (2006). Sinonasal undifferentiated carcinoma. Am J ClinOncol, 29, 27-31.
Odell EW, Jani P, Sherriff M, et al. (1994). The prognostic value of individual grading parameters in small lingual squamous cell carcinomas. The importance of the pattern of invasion. Cancer, 74, 789–94.
Oosterkamp S, de Jong JM, Van den Ende PL, et al. (2006). Predictive value of lymph node metastases and extracapsular extension for the risk of distant metastases in laryngeal carcinoma.Laryngoscope, 116(11), 2067-70.
Pentenero M, Gandolfo S, Carrozzo M. (2005). Importance of tumor thickness and depth of invasion in nodal involvement and prognosis of oral squamous cell carcinoma: a review of the literature. Head Neck, 27, 1080–91.
Puri SK, Fan CY, Hanna E. (2003). Significance of extracapsular lymph node metastases in patients with head and neck squamous cell carcinoma.CurrOpinOtolaryngol Head Neck Surg, 11(2), 119-23.
Ragin CC, Taioli E. (2007). Survival of squamous cell carcinoma of the head and neck in relation to human papillomavirus infection: review and meta-analysis. Int J Cancer, 121, 1813–20.
Regis De Brito Santos I, Kowalski LP, Cavalcante De Araujo V, et al. (2001). Multivariate analysis of risk factors for neck metastases in surgically treated parotid carcinomas. Arch Otolaryngol Head Neck Surg, 127, 56–60.

Renner G. (2007). Small cell carcinoma of the head and neck: a review. SeminOncol, 34, 3-14.
Robinson M, Schache A, Sloan P, et al. (2012). HPV specific testing: a requirement for oropharyngeal squamous cell carcinoma patients. Head Neck Pathol., 6(Suppl1), S83-90.
Schilling C, Stoeckli SJ, Vigili MG, de Bree R, Lai SY, Alvarez J, Christensen A, Cognetti DM, D'Cruz AK, Frerich B, Garrel R, Kohno N, Klop WM, Kerawala C, Lawson G, McMahon J, Sassoon I, Shaw RJ, Tvedskov JF, von Buchwald C, McGurk M. (2019). Surgical consensus guidelines on sentinel node biopsy (SNB) in patients with oral cancer.Head Neck. 41(8), 2655-2664.
Seethala RR. (2011). Histologic grading and prognostic biomarkers in salivary gland carcinomas.AdvAnatPathol, 18, 29–45.
Sloan P. (2009). Head and neck sentinel lymph node biopsy: current state of the art. Head Neck Pathol, 3(3), 231-7.
Slootweg PJ, Ferlito A, Cardesa A, et al. (2013). Sinonasal tumors: a clinicopathologic update of selected tumors. Eur Arch Otorhinolaryngol, 270(1), 5-20.
Smeets SJ, Hesselink AT, Speel EJ, et al. (2007). A novel algorithm for reliable detection of human papillomavirus in paraffin embedded head and neck cancer specimen. Int J Cancer,121(11), 2465-72.
Snow GB, Annayas A, van Slooten EA, et al. (1982). Prognostic factors of neck node metastasis. ClinOtolaryngol, 7, 185–190.
Speight PM, Barrett AW. (2009). Prognostic factors in malignant tumours of the salivary glands. Br J Oral MaxillofacSurg, 47, 587–593.
Stelow EB, Bellizzi AM, Taneja K, et al. (2008). NUT rearrangement in undifferentiated carcinomas of the upper aerodigestive tract. Am J SurgPathol, 32, 828–834.
Stelow EB, Mills SE, Jo VY, et al. (2010). Adenocarcinoma of the upper aerodigestive tract. AdvAnatPathol, 17, 262–269.
Terhaard CH, Lubsen H, Van der Tweel I, et al. (2004). Salivary gland carcinoma: independent prognostic factors for locoregional control, distant metastases, and overall survival: results of the Dutch head and neck oncology cooperative group. Head Neck, 26, 681–692.
Thompson LD. (2009). Olfactory neuroblastoma. Head Neck Pathol, 3,252-9.
Trivedi NP, Ravindran HK, Sundram S, et al. (2010). Pathologic evaluation of sentinel lymph nodes in oral squamous cell carcinoma. Head Neck. 32(11), 1437-43.
Van der Schroeff MP, Terhaard CH, Wieringa MH, et al. (2010). Cytology and histology have limited added value in prognostic models for salivary gland carcinomas. Oral Oncol,46, 662–666.
Van Gompel JJ, Giannini C, Olsen KD, et al. (2012). Long-term outcome of esthesioneuroblastoma: hyams grade predicts patient survival. J NeurolSurg B Skull Base, 73, 331-6.
Wan XC, Egloff AM, Johnson J. (2012). Histological assessment of cervical lymph node identifies patients with head and neck squamous cell carcinoma (HNSCC): who would benefit from chemoradiation aftersurgery? Laryngoscope, 122(12), 2712-22.
Weijers M, Snow GB, Bezemer PD, et al. (2009). Malignancy grading is no better than conventional histopathological grading in small squamous cell carcinoma of tongue and floor of mouth: retrospective study in 128 patients. J Oral Pathol Med., 38(4), 343-7.
Westra WH. (2012). Detection of human papillomavirus in clinical samples. OtolaryngolClin North Am., 45(4), 765-77.

Woolgar JA, Rogers S, West CR, et al. (1999). Survival and patterns of recurrence in 200 oral cancer patients treated byradical surgery and neck dissection. Oral Oncol., 35(3), 257-65

Woolgar JA, Triantafyllou A, Lewis JS Jr, et al. (2013). Prognostic biological features in neck dissection specimens. Eur Arch Otorhinolaryngol, 270(5), 1581-92.

Woolgar JA, Triantafyllou A. (2005). A histopathological appraisal of surgical margins in oral and oropharyngeal cancer resection specimens. Oral Oncol., 41(10), 1034-43.

Woolgar JA, Triantafyllou A. (2009). Pitfalls and procedures in the histopathological diagnosis of oral and oropharyngeal squamous cell carcinoma and a review of the role of pathology in prognosis. Oral Oncol., 45(4-5), 361-85.

Woolgar JA. (2006). Histopathological prognosticators in oral and oropharyngeal squamous cellcarcinoma. Oral Oncol, 42, 229–39.

Geldigheid en Onderhoud

Module[1]	Regiehouder(s)[2]	Jaar van autorisatie	Eerstvolgende beoordeling actualiteit richtlijn[3]	Frequentie van beoordeling op actualiteit[4]	Wie houdt er toezicht op actualiteit[5]	Relevante factoren voor wijzigingen in aanbeveling[6]

[1] Naam van de module

[2] Regiehouder van de module (deze kan verschillen per module en kan ook verdeeld zijn over meerdere regiehouders)

[3] Maximaal na vijf jaar

[4] (half)Jaarlijks, eens in twee jaar, eens in vijf jaar

[5] regievoerende vereniging, gedeelde regievoerende verenigingen, of (multidisciplinaire) werkgroep die in stand blijft

[6] Lopend onderzoek, wijzigingen in vergoeding/organisatie, beschikbaarheid nieuwe middelen

Bijlagen bij module 10.1

1. (6.1)

Gradering mucoepidermoid carcinoom

Gradering mucoepidermoid carcinoom (AFIP/WHO)

Intracysteuze component <20%	2
Neurale invasie	2
Necrose	3
Mitosen (4 of meer per 10 HPF)	3
Anaplasie	4

Low grade = 0–4 pts
Intermediate grade = 5–6 pts
High grade = 7–14 pts

2. (6.2)

Gradering esthesioneuroblastoom

Esthesioneuroblastoma grading (based on Hyams' grading system)

Microscopic features	Grade I	Grade II	Grade III	Grade IV
Architecture	Lobular	Lobular	±Lobular	±Lobular
Pleomorphism	Absent to slight	Present	Prominent	Marked
NF matrix	Prominent	Present	May be present	Present
Rosettes	HR	HR	FW	FW
Mitoses	Absent	Present	Prominent	Marked
Necrosis	Absent	Absent	Present	Prominent
Glands	May be present	May be present	May be present	May be present
Calcification	Variable	Variable	Absent	Absent

NF neurofibrillary, HR Homer Wright pseudorosettes, FW Flexner-Wintersteiner rosettes

Kennislacunes

Geen.

Implementatieplan

Aanbeveling	Tijdspad voor implementatie: < 1 jaar, 1 tot 3 jaar of 3 tot 5 jaar	Verwacht effect op kosten	Randvoorwaarden voor implementatie (binnen aangegeven tijdspad)	Mogelijke barrières voor implementatie[1]	Te ondernemen acties voor implementatie[2]	Verantwoordelijken voor acties[3]	Overige opmerkingen
Vermeld op het aanvraagformulier voor algemeen pathologie-onderzoek het volgende: De clinicus vermeldt **altijd** de volgende gegevens: • ingreep; • tumorlokalisatie; • zijdigheid • klinische TN classificatie;	< 1 jaar	Geen	Geen	Geen	Opname in de PALGA protocol module, waarvoor overleg met Stichting PALGA nodig is	NVVP	Geen

• relevante voorgeschiedenis (eerdere behandeling, relevante comorbiditeit). De clinicus vermeldt bij halsklierdissectie **aanvullend** de volgende gegevens: • type halsklierdissectie; • levels van het halsklierdissectiepreparaat; • markering van levels op preparaat of bijgevoegde tekening. De clinicus vermeldt bij een schildwachtklierprocedure **aanvullend** de volgende gegevens: welke klier(en) volgens de schildwachtklierprocedure bewerkt moeten worden.							
Vermeld in het verslag pathologie-onderzoek onderstaande en **gebruik** hierbij bij voorkeur de PALGA protocol module: De patholoog vermeldt minimaal de volgende algemene gegevens: • naam aanvrager; • datum; • data patiënt; • klinische gegevens (letterlijk overgenomen van het aanvraagformulier); • naam patholoog;	< 1 jaar	Geen	Geen	Geen	Opname in de PALGA protocol module, waarvoor overleg met Stichting PALGA nodig is	NVVP	Geen

• macroscopie; • microscopie; • conclusie. **Specifieke** gegevens verslag *mucosaal plaveiselcelcarcinoom* Het verslag van een *biopt* bevat minimaal de volgende gegevens: • histologisch type; • differentiatiegraad optioneel; • bij orofarynxbiopt: HPV bepaling. Het verslag van een *resectiepreparaat* bevat minimaal de volgende gegevens: • aard preparaat, zijdigheid; • plaats tumor; • histologisch type en differentiatiegraad tumor; • groeipatroon; • maximale diameter; • invasiediepte; • kleinste afstand carcinoom tot de resectievlakken in mm, zowel mucosale- als weke delen resectievlakken; • aan- of afwezigheid (lymf)vaat-ingroei; • aan- of afwezigheid perineurale groei ter plaatse van het invasieve front; • aan- of afwezigheid (ernstige) dysplasie in de							

mucosale resectievlakken; • indien van toepassing: • aan- of afwezigheid kraakbeen- of botinvasie; • aan- of afwezigheid doorgroei in de schildklier. • aan- of afwezigheid humaan papillomavirus (HPV) bij plaveiselcelcarcinomen van de orofarynx. **Specifieke** gegevens verslag *speekselkliercarcinomen* Het verslag van een *resectiepreparaat* bevat minimaal de volgende gegevens: • aard preparaat, zijdigheid; • histologisch tumortype en indien relevant (bijvoorbeeld mucoepidermoïd carcinoom) gradering; • maximale diameter; • ingroei in omgevende structuren; • perineurale groei in “named nerves”; • kleinste afstand carcinoom tot de resectievlakken in mm **Specifieke** gegevens verslag *neus- en neusbijholtetumoren* Het verslag van een *resectiepreparaat* bevat minimaal de							

volgende gegevens: • aard preparaat, zijdigheid; • histologisch tumortype en indien relevant (bijvoorbeeld esthesioneuroblastoom) gradering; • maximale diameter; • ingroei in omgevende structuren (bot); • perineurale groei in “named nerves”; • Kleinste afstand carcinoom tot de resectievlakken in mm, zowel mucosale- als weke delen resectievlakken. **Specifieke** gegevens verslag *lymfeklierresectie* Het verslag van een *lymfeklierresectiepreparaat* van de hals bij een primair hoofd-hals carcinoom of bij een metastase ven onbekende primaire tumor bevat minimaal de volgende gegevens: • aard preparaat (type halsklierdissectie met vermelding levels); • per level: • aantal lymfklieren; • aantal lymfklieren met metastasen; • diameter van de grootste metastase; • aan- of afwezigheid van							

extranodale groei. **Specifieke** gegevens verslag *schildwachtklierprocedure* Het verslag van een *schildwachtklierprocedure* bevat minimaal de volgende gegevens: • aantal lymfklieren; • aan- of afwezigheid van een metastase; • diameter van de metastase (Isolated Tumor Cells/micro-/macrometastasen); • aan- of afwezigheid van extranodale groei.							

[1] Barrières kunnen zich bevinden op het niveau van de professional, op het niveau van de organisatie (het ziekenhuis) of op het niveau van het systeem (buiten het ziekenhuis). Denk bijvoorbeeld aan onenigheid in het land met betrekking tot de aanbeveling, onvoldoende motivatie of kennis bij de specialist, onvoldoende faciliteiten of personeel, nodige concentratie van zorg, kosten, slechte samenwerking tussen disciplines, nodige taakherschikking, et cetera.
[2] Denk aan acties die noodzakelijk zijn voor implementatie, maar ook acties die mogelijk zijn om de implementatie te bevorderen. Denk bijvoorbeeld aan controleren aanbeveling tijdens kwaliteitsvisitatie, publicatie van de richtlijn, ontwikkelen van implementatietools, informeren van ziekenhuisbestuurders, regelen van goede vergoeding voor een bepaald type behandeling, maken van samenwerkingsafspraken.
[3] Wie de verantwoordelijkheden draagt voor implementatie van de aanbevelingen, zal tevens afhankelijk zijn van het niveau waarop zich barrières bevinden. Barrières op het niveau van de professional zullen vaak opgelost moeten worden door de beroepsvereniging. Barrières op het niveau van de organisatie zullen vaak onder verantwoordelijkheid van de ziekenhuisbestuurders vallen. Bij het oplossen van barrières op het niveau van het systeem zijn ook andere partijen, zoals de NZA en zorgverzekeraars, van belang.

Module 12.1 Behandeling T1-2N0-1 orofarynxcarcinomen

Uitgangsvraag
Hoe worden patiënten met T1-2N0-1 orofaryngeale tumoren behandeld?

Inleiding
De standaardbehandeling van patiënten met T1-2N0-1 orofarynxcarcinomen is radiotherapie. Transorale robotische chirurgie (Transoral Robotic Surgery, TORS) wordt in aantal ziekenhuizen aangeboden als alternatief. In de Verenigde Staten is TORS geregistreerd voor deze indicatie door de Food and Drugs Administration (FDA). Gezien het feit dat deze aandoening bij een steeds jongere patientenpopulatie vastgesteld wordt door de toename van HPV+ tumoren en gelet de toxiciteit van de radiotherapie, zeker op lange termijn, wordt de behandeling middels TORS al dan niet in combinatie met (een eventueel lagere dosis) radiotherapie intensief onderzocht.

Search and select
A systematic review of the literature was performed to answer the following question:

> What are the (un)beneficial effects of TORS (with or without transoral microsurgery, and with or without adjuvant radiotherapy) on overall survival, disease-free survival, local control and quality of life, e.g. swallowing, trismus, taste, dryness of the mouth, and mucositis in patients with a T1-2N0-1 oropharyngeal carcinoma, when compared to primary radiotherapy.

P: patients with a T1-2N0-1 oropharyngeal carcinoma;
I: transoral robotic surgery with or without transoral microsurgery and with or without adjuvant radiotherapy;
C: primary radiotherapy;
O: overall survival, disease-free survival, local control, swallowing (complaints), trismus, taste, dryness of the mouth, mucositis, quality of life.

<u>Relevant outcome measures</u>
The guideline development group considered overall survival, disease-free survival, local control, swallowing, and quality of life (as measured with the MD Anderson Dysphagia Index (MDADI)) as a critical outcome measure for decision making; and trismus, dryness of the mouth, and mucositis as an important outcome measure for decision making.

A priori, the working group did not define the outcome measures listed above but used the definitions used in the studies.

The working group defined a minimal clinically relevant difference as *(in line with "NVMO-commissie ter Beoordeling van Oncologische Middelen (BOM)")*:
- > 5% difference or more >3% and HR <0.7 in overall survival.
- HR< 0.7 for progression free survival.

And, in case of absence of a clinically relevant difference in overall survival or progression free survival:
- A minimal clinically important difference of 10 points on the quality of life instrument EORTC QLQ-C30 (in line with Mehanna, 2019) or a difference of a similar magnitude on other quality of life instruments.
- Statistically significant less complications/adverse events.
- Statistically significant better functional outcomes, work participation.

Search and select (Methods)
The databases Medline (via OVID) and Embase (via Embase.com) were searched with relevant search terms until 6th of January 2019 for systematic reviews and randomized controlled trials. The detailed search strategy is depicted under the tab Methods. The systematic literature search resulted in 106 hits. Studies or systematic reviews were selected based on the following criteria: patients had a T1-2N0-1 oropharyngeal carcinoma, transoral robotic surgery was compared with primary radiotherapy, at least one of the outcomes of interest was reported (overall survival, disease-free survival, local control, swallowing (complaints), trismus, dryness of the mouth, mucositis, quality of life), the study design was a randomized controlled trial or the systematic review contained randomized controlled trials, reports had to be written in English or Dutch. Twenty-one studies were initially selected based on title and abstract screening. After reading the full text, 20 studies were excluded (see the table with reasons for exclusion under the tab Evidence tables) and 1 randomized controlled trial was included.

Results
One randomized controlled trial (Nichols, 2019) was included in the analysis of the literature. Important study characteristics and results are summarized in the evidence tables. The assessment of the risk of bias is summarized in the risk of bias tables (under the tab Evidence tables).

Summary of literature
Description of studies
Transoral robotic surgery (TORS) versus primary radiotherapy (RT) in Nichols (2019)
Nichols (2019) reported on a multicenter randomized controlled trial with centers recruiting in Canada and Australia to compare TORS with neck dissection (on indication) with primary radiotherapy (RT) as part of the ORATOR-trial. Data and procedures were extracted from the study protocol (Nichols, 2013) and the study report with appendices (Nichols, 2019).

Nichols (2019) included patients who were over 18 years old, had an ECOG-status of 0 to 2, had a histologically confirmed squamous cell carcinoma where the primary site was the oropharynx, had a T1-2 carcinoma and were likely to have negative resection margins at surgery, had N0-2 without extranodal extension (on pre-randomization imaging), the complete blood count/differential was obtained 4 week prior to randomization (with adequate bone marrow, hepatic, and renal function), and the patient was assessed at a multidisciplinary head and neck clinic and presented at a multidisciplinary tumor board prior to randomization. Patients were excluded when there were serious medical comorbidities, there were contraindications to therapy (radiotherapy, chemotherapy, surgery), there was prior history of head and neck cancer within 5 years, the patient had prior head or neck radiation, the patient had metastases, the patient was unable to attend full-course radiotherapy or follow-up visits, the patient had prior malignant disease (unless 5-year disease-free) with the exception of non-melanoma skin cancer, the patient was pregnant or lactating, or when the patient was unwilling or unable to complete the quality of life questionnaires.

A total of 68 patients were recruited and randomly allocated (1:1 allocation, block size 4). Patients in the TORS-group (n=34, 28 males) had a median age of 58.2 years (IQR: 52.6 to 64.5) and the primary tumor located at the tonsil / tonsillar fossa (n=24) or at the base of the tongue (n=10). Patients had an ECOG score of either 0 (n=30) or 1 (n=4), while 6 out of 22 (27%) who were asked drank over 21 drinks per week. Twenty-one (62%) patients undergoing TORS had a smoking history. Clinical T-stage was either T1 (n=17) or T2 (n=17),

while clinical N-stage was N0 (n=9), N1 (n=7), or N2 (n=18). There were 30 HPV-positive patients in the TORS-group, as measured with p16 staining. Patients in the radiotherapy-group (n=34, 31 males) had a median age of 60 years (IQR: 53.2 to 65.2) and the primary tumor located at the tonsil / tonsillar fossa (n=26) or at the base of the tongue (n=8). Patients had an ECOG score of 0 (n=30) or 1 (n=4). Eighteen persons were asked about the alcohol-intake, resulting in 1 person (6%) having more than 21 drinks per week. Twenty-eight (82%) patients had a smoking history. In the RT group, the clinical T-stage was T1 (n=13) or T2 (n=21), while the clinical N-stage was N0 (n=12), N1 (n=5), or N2 (n=17). There were 30 HPV-positive patients by p16 staining in the RT-group as well.

In the TORS-group, a surgical robot was used to excise the primary oropharyngeal tumor with the spatula caultry (1cm resection margin). Selective neck dissection was performed at the discretion of the surgeon during surgery or within 2 weeks after surgery. Surgeons had to perform at least 10 TORS excisions prior to enrolling patients in the study. The surgical specimen was sent for frozen section analysis and the excision continued until negative margins were obtained. Adjuvant radiotherapy was given, however could be omitted when there was no extranodal extension, positive margins, pT3-4, nodal disease, or lymphovascular invasion. Thus, radiotherapy was given in the region of positive margins (64 Gy, 30 fractions, 6 weeks), in high risk nodal areas (60 Gy, 30 fractions, 6 weeks), or in low risk nodal areas (54 Gy, 30 fractions, 6 weeks). Concurrent chemotherapy with cisplatin (100mg/m^2 every 3 weeks in a 3 week cycle) could be administered when there were positive margins or extracapsular extensions. When patients were deemed unfit for cisplatin chemotherapy the doses and schedules could be modified, or cetuximab or weekly carboplatin (AUC 1.5) could be administered at the discretion of the medical oncologist. Out of the 34 patients in the TORS-group, 24 (71%) patients received radiotherapy and 8 (24%) patients received chemotherapy (cisplatin: n=5, carboplatin: n=3). Patients received chemotherapy in a median number of cycles of 6 (IQR: 4.5 to 6). It was unclear whether there were any deviations from the cisplatin schedule (100mg/m^2 every 3 weeks in a 3 week cycle) at the medical oncologist's discretion.

In the RT-group, primary radiotherapy was given to the gross tumor and nodes (70 Gy, 35 fractions, 7 weeks). Radiotherapy was further administered in high risk nodal areas (63 Gy, 35 fractions, 7 weeks) and in low risk nodal areas (56 Gy, 35 fractions, 7 weeks). At the treating radiation oncologist's discretion the radiotherapy regimen could be accelerated or hyperfractionated, where the same dose could be provided in 6 weeks. Concurrent chemotherapy was administered with positive nodes (i.e. N1-2) and omitted when N0. Chemotherapy doses and schedules could be modified for patients who were deemed unfit for cisplatin, or cetuximab or weekly carboplatin (AUC 1.5) could be administered at the discretion of the medical oncologist. Twenty-three (68%) patients received chemotherapy (cisplatin: n=19, carboplatin: n=3, cetuximab: n=1). Patients received chemotherapy in a median number of cycles of 3 (IQR: 2-6). It was unclear whether there were any deviations from the cisplatin schedule (100mg/m2 every 3 weeks in a 3 week cycle) at the medical oncologist's discretion.Treatment response was evaluated by CT or PET-CT 8 to 12 weeks after completion of radiotherapy. When the CT showed residual nodes over 1 centimeter, patients were recommended salvage surgery. For PET-CT evaluation, residual nodes had to be over 1 centimeter and FDG-avid. Patients were offered surgical salvage when feasible for relapse or progressive disease. Four (12%) patients received salvage surgery, however it was unclear whether nodes were examined by a pathologist to confirm metastases.

Patients were followed-up every 3 months in the first two years and every 6 months in year 3 to 5. Physical examination and adverse event monitoring took place at every visit. Quality

of life measurements and a chest x-ray was performed every 6 months, with the exception of the 1-year time point where a CT-scan was taken instead of an x-ray. The RT-group had an additional assessment at 8 to 12 weeks post-radiotherapy to assess the treatment response and the need for salvage surgery. The median follow-up in de study was 27 months (IQR: 20 to 48). Two patients in the RT-group withdrew consent after randomization (no reasons provided).

Nichols (2019) measured quality of life (QoL) with several validated questionnaires: MD Anderson Dysphagia Index (MDADI, score range: 20 to 100, higher indicates a better QoL), European Organization for Research and Treatment of Cancer Quality of Life of Cancer Patients general (EORTC QLQ-C30, score range: 0 to 100, higher score indicates a worse QoL), Organization for Research and Treatment of Cancer Quality of Life of Cancer Patients Head and Neck (EORTC H&N35, score range: 0 to 100, higher score indicates a worse QoL), Voice Handicap Index (VHI-10, score range: 0 to 40, higher score indicates a better QoL), Neck Dissection Impairment Index (NDII, score range: 0 to 100, higher score indicates a better QoL), Patient Neurotoxicity Questionnaire (PNQ, 2 items on a 5-point Likert scale), and Functional Oral Intake Scale (FOIS, 7-point Likert scale, higher scores indicate a better QoL).

Results

Transoral robotic surgery (TORS) versus primary radiotherapy (RT)

Overall survival

Nichols (2019) reported the 5-year overall survival of both groups, each containing 34 patients (reference group=TORS; log-rank p=0.89; HR=0.83, 95%CI: 0.21 to 8.35). Over the course of 5 years, there were 4 observed deaths in the RT-group (n=3 metastatic disease, n=1 cardiac event outside the hospital) and 5 observed deaths in the TORS group (n=3 metastatic disease, n=1 TORS-related bleeding, n=1 cardiac arrest after alcohol overdose and hypoglycaemia). A total of 27 persons in the TORS-group and 28 persons in the RT-group were censored over the course of 5 years. No reasons for censoring were provided.

Disease-free survival

Nichols (2019) defined the 5-year 'progression-free survival' as the time from randomization to either death or recurrence, whichever occurred first. A hazard ratio of 1.07 (95%CI: 0.28 to 4.01; reference group=TORS) and a log-rank p of 0.63 were reported. A total of 26 persons in the TORS-group and 28 persons in the RT-group were censored over the course of 5 years. No reasons for censoring were provided. There were 4 recurrences observed in each group.

Swallowing

Nichols (2019) reported swallowing complications per grade. The relative risk (RR) was not reported, however from these data the RR could be calculated. For grade 1 to 2 dysphagia an RR of 1.00 (95%CI: 0.69 to 1.45, reference group=RT) was reported. Both in the TORS and RT-group 21 grade 1-2 events were observed. Grade 3 dysphagia had 9 events in the TORS-group and 6 events in the RT-group. Here, an RR of 1.50 (95%CI: 0.60 to 3.75) was calculated. When all observed events were combined (TORS: 30 events, RT: 27 events) the RR was 1.11 (95%CI: 0.90 to 1.37). One person in the RT-group used a percutaneous feeding tube at the 1 year time point, compared to none in the TORS-group.

Local control

No studies were included that reported the local control.

Quality of life

Nichols (2019) assessed the 1-year quality of life with several questionnaires, however the quality of life as measured with the MDADI was the primary outcome measure. The MDADI total score differed significantly between groups at 1 year (TORS: 80.1 (SD: 13), RT: 86.9 (SD: 11.4), p=0.042), indicating the RT-group had a better swallowing-related quality of life. Similarly, a statistically significant difference was reported for the MDADI composite score (TORS: 80.2 (SD: 13.1), RT: 86.7 (SD: 11.4), p=0.049), indicating a better swallowing-related quality of life for the RT-group. The total score on the VHI-10 (TORS: 4.5 (SD: 4.3), RT: 4.4 (SD: 4.6), p=0.89) and NDII (TORS: 81.5 (SD: 28.7), RT: 92.3 (SD: 10), p=0.072) showed no significant difference. Summary scores for the EORTC QLQ-C30 and the EORTC H&N35 were not reported. For scores on sub-scales of the questionnaires we refer to Table 12.1.

Table 12.1 Overview of the 1-year quality of Life measurements per questionnaire as measured by Nichols (2019)

Quality of life instrument	Outcome
MD Anderson Dysphagia Index (MDADI, score range: 20-100, higher indicates a better QoL), completed surveys at the 1-year time point: 30 for TORS, 27 for RT.	<u>MDADI total score, mean score (SD):</u> TORS: 80.1 (13) RT: 86.9 (11.4) Mean difference: 6.7 (95%CI: 0.2 to 13.2) Statistically significant difference between groups: p=0.042 <u>MDADI global sub-scale score, mean score (SD):</u> TORS: 79.3 (22.6) RT: 89.6 (15.1) Mean difference: 10.3 (95%CI: 0.2 to 20.4) Statistically significant difference between groups: p=0.046 <u>MDADI emotional sub-scale score, mean score (SD):</u> TORS: 81.3 (12.5) RT: 88.8 (12) Mean difference: 7.4 (95%CI: 0.9 to 14) Statistically significant difference between groups: p=0.027 <u>MDADI functional sub-scale score, mean score (SD):</u> TORS: 86.5 (12) RT: 89.9 (11.5) Mean difference: 3.4 (95%CI: -2.9 to 9.6) No statistically significant difference between groups: p=0.28 <u>MDADI physical sub-scale score, mean score (SD):</u> TORS: 75.3 (16.5) RT: 83.1 (14.1) Mean difference: 7.9 (95%CI: -0.3 to 16) No statistically significant difference between groups: p=0.058 <u>MDADI composite score, mean score (SD):</u> TORS: 80.2 (13.1) RT: 86.7 (11.4) Mean difference: 6.5 (95%CI: 0.0 to 13.1) Statistically significant difference between groups: p=0.049
European Organization for Research and Treatment of Cancer Quality of Life of Cancer Patients General (EORTC QLQ-C30, score range: 0-100, higher score indicates a worse QoL on symptom scales, a high score on functional and global health scales represent higher functioning/QoL), unclear number of completed surveys at the 1-year time point.	<u>EORTC-C30 Global health status sub-scale score, mean score (SD):</u> TORS: 77.9 (19.5) RT: 76.2 (20.9) No statistically significant difference between groups: p= 0.76 <u>EORTC-C30 Physical functioning sub-scale score, mean score (SD):</u> TORS: 9.4 (16.1) RT: 5.9 (7.2)

	No statistically significant difference between groups: p= 0.29 <u>EORTC-C30 Role functioning sub-scale score, mean score (SD):</u> TORS: 18.3 (30.1) RT: 11.1 (17.9) No statistically significant difference between groups: p= 0.27 <u>EORTC-C30 Emotional functioning sub-scale score, mean score (SD):</u> TORS: 14.9 (19.5) RT: 12.0 (15.9) No statistically significant difference between groups: p= 0.54 <u>EORTC-C30 Cognitive functioning sub-scale score, mean score (SD):</u> TORS: 13.8 (18.9) RT: 11.7 (15.9) No statistically significant difference between groups: p=0.66 <u>EORTC-C30 Social functioning status sub-scale score, mean score (SD):</u> TORS: 13.2 (20.1) RT: 6.4 (13.4) No statistically significant difference between groups: p= 0.14 <u>EORTC-C30 Fatigue sub-scale score, mean score (SD):</u> TORS: 18.1 (20.5) RT: 15.6 (13.5) No statistically significant difference between groups: p= 0.59 <u>EORTC-C30 Nausea/ vomiting sub-scale score, mean score (SD):</u> TORS: 5.0 (9.9) RT: 4.3 (10.9) No statistically significant difference between groups: p= 0.81 <u>EORTC-C30 Pain sub-scale score, mean score (SD):</u> TORS: 21.8 (25.2) RT: 8.0 (16.3) Statistically significant difference between groups: p= 0.018 <u>EORTC-C30 Dyspnea sub-scale score, mean score (SD):</u> TORS: 7.8 (14.3) RT: 4.9 (12.1) No statistically significant difference between groups: p= 0.42 <u>EORTC-C30 Insomnia sub-scale score, mean score (SD):</u> TORS: 17.8 (27.3) RT: 28.4 (28.8) No statistically significant difference between groups: p= 0.16 <u>EORTC-C30 Appetite loss sub-scale score, mean score (SD):</u> TORS: 13.3 (27.1) RT: 16.0 (25.1) No statistically significant difference between groups: p= 0.70 <u>EORTC-C30 Constipation sub-scale score, mean score (SD):</u> TORS: 4.4 (14.5) RT: 8.6 (14.9) No statistically significant difference between groups: p= 0.29 <u>EORTC-C30 Diarrhea sub-scale score, mean score (SD):</u> TORS: 5.7 (15.6)

	RT: 2.5 (8.9) No statistically significant difference between groups: p= 0.34 <u>EORTC-C30 Financial difficulties sub-scale score, mean score (SD):</u> TORS: 14.9 (29.0) RT: 11.1 (24.5) No statistically significant difference between groups: p= 0.59
Organization for Research and Treatment of Cancer Quality of Life of Cancer Patients Head and Neck (EORTC H&N35, score range: 0-100, higher score indicates a worse QoL on symptom scales, a high score on functional scales represent higher functioning/QoL), unclear number of completed surveys at the 1-year time point.	<u>EORTC QLQ-HN35 Pain sub-scale score, mean score (SD):</u> TORS: 13.3 (14.9) RT: 9.0 (12.4) No statistically significant difference between groups: p= 0.23 <u>EORTC QLQ-HN35 Swallowing sub-scale score, mean score (SD):</u> TORS: 12.7 (16.1) RT: 7.4 (7.4) No statistically significant difference between groups: p= 0.11 <u>EORTC QLQ -HN35 Senses sub-scale score, mean score (SD):</u> TORS: 20.6 (21.3) RT: 20.5 (22.8) No statistically significant difference between groups: p > 0.99 <u>EORTC QLQ -HN35 Speech sub-scale score, mean score (SD):</u> TORS: 7.7 (9.9) RT: 5.8 (9.9) No statistically significant difference between groups: p= 0.48 <u>EORTC QLQ -HN35 Social eating sub-scale score, mean score (SD):</u> TORS: 11.8 (14.4) RT: 7.1 (10.3) No statistically significant difference between groups: p= 0.16 <u>EORTC QLQ -HN35 Social contact sub-scale score, mean score (SD):</u> TORS: 4.6 (10.1) RT: 1.5 (6.5) No statistically significant difference between groups: p= 0.17 <u>EORTC QLQ -HN35 Less sexuality sub-scale score, mean score (SD):</u> TORS: 22.0 (27.6) RT: 17.3 (25.2) No statistically significant difference between groups: p= 0.51 <u>EORTC QLQ -HN35 Teeth sub-scale score, mean score (SD):</u> TORS: 12.2 (22.3) RT: 1.2 (6.4) Statistically significant difference between groups: p= 0.014 <u>EORTC QLQ -HN35 Opening mouth sub-scale score, mean score (SD):</u> TORS: 11.1 (22.0) RT: 6.4 (13.4) No statistically significant difference between groups: p= 0.33 <u>EORTC QLQ -HN35 Dry mouth sub-scale score, mean score (SD):</u> TORS: 44.4 (30.7) RT: 53.1 (31.0) No statistically significant difference between groups: p= 0.30

	EORTC QLQ -HN35 Sticky saliva sub-scale score, mean score (SD): TORS: 31.1 (34.9) RT: 32.1 (28.5) No statistically significant difference between groups: p= 0.91 EORTC QLQ -HN35 Coughing sub-scale score, mean score (SD): TORS: 22.2 (23.7) RT: 24.7 (25.5) No statistically significant difference between groups: p= 0.71 EORTC QLQ -HN35 Felt ill sub-scale score, mean score (SD): TORS: 6.7 (13.6) RT: 3.7 (10.7) No statistically significant difference between groups: p= 0.36 EORTC QLQ -HN35 Pain killers sub-scale score, mean score (SD): TORS: 44.8 (50.6) RT: 14.8 (36.2) Statistically significant difference between groups: p= 0.013 EORTC QLQ -HN35 Nutritional supplements sub-scale score, mean score (SD): TORS: 24.1 (43.5) RT: 29.6 (46.5) No statistically significant difference between groups: p= 0.65 EORTC QLQ -HN35 feeding tube sub-scale score, mean score (SD): TORS: 0.0 (0) RT: 3.7 (19.2) No statistically significant difference between groups: p= 0.33 EORTC QLQ -HN35 weight loss sub-scale score, mean score (SD): TORS: 20.7 (41.2) RT: 3.7 (19.2) No statistically significant difference between groups: p= 0.053 EORTC QLQ -HN35 weight gain sub-scale score, mean score (SD): TORS: 37.9 (49.4) RT: 40.7 (50.1) No statistically significant difference between groups: p= 0.83
Voice Handicap Index (VHI-10, score range: 0-40, higher score indicates a better QoL), unclear number of completed surveys at the 1-year time point.	VHI-10 total score, mean score (SD): TORS: 4.5 (4.3) RT: 4.4 (4.6) No statistically significant difference between groups: p= 0.89
Neck Dissection Impairment Index (NDII, score range: 0-100, higher score indicates a better QoL), unclear number of completed surveys at the 1-year time point.	NDII total score, mean score (SD): TORS: 81.5 (28.7) RT: 92.3 (10.0) No statistically significant difference between groups: p= 0.072
Patient Neurotoxicity Questionnaire (PNQ, 2 items on a 5-point Likert scale), unclear number of completed surveys at the 1-year time point.	PNQ Numbness sub-scale score, mean score (SD): TORS: 0.5 (0.8) RT: 0.3 (0.6) No statistically significant difference between groups: p= 0.57 PNQ Weakness sub-scale score, mean score (SD):

	TORS: 0.5 (1.0) RT: 0.2 (0.5) No statistically significant difference between groups: p= 0.14
Functional Oral Intake Scale (FOIS, 7-point Likert scale, higher scores indicate a better QoL), unclear number of completed surveys at the 1-year time point.	FOIS, level 7 "expansion of oral diet" reached at 1 year, n(%): TORS: 26/31 (84%) RT: 27/27 (100%) No statistically significant difference between groups: p= 0.055
QoL: Quality of Life **RT: Radiotherapy-group** **TORS: Transoral robotic surgery-group**	

Trismus
Nichols (2019) reported grade 1-2 trismus in 8 of the 34 patients (23.5%) in the TORS-group and in 1 of the 34 patients (2.9%) in the RT-group. The RR was not reported, however from these data the RR could be calculated. Based on the data an RR of 8.00 (95%CI: 1.06 to 60.5) was calculated. One patient in the TORS-group had a grade 3 trismus complication, therefore the overall RR was 9.00 (95%CI: 1.21 to 67.21).

Taste
Grade 1 to 2 taste alterations were observed by Nichols (2019) in 17 patients (50%) in the TORS-group, compared to 19 patients (55.9%) in the RT group. The RR was not reported, however from these data the RR could be calculated. An RR of 0.89 (95%CI: 0.57 to 1.40) was calculated. Taste alterations in grade 3 or higher were not prevalent in both groups.

Dryness of the mouth
Grade 1 to 2 dryness of the mouth was reported in 18 of the 34 patients (52.9%) in the TORS-group, compared to 24 out of 34 patients (70.6%) in the RT-group. The RR was not reported, however from these data the RR could be calculated. The calculated RR was 0.75 (95%CI: 0.51 to 1.10). One patient in the RT-group had a grade 3 dryness of the mouth. The RR of all observed events was 0.72 (95%CI: 0.49 to 1.05, TORS: 18 events, RT: 25 events).

Mucositis
Nichols (2019) reported oral mucositis in 8 (23.5%) patients in the TORS-group, compared to 11 (32.4%) patients in the RT-group. The RR was not reported, however from these data the RR could be calculated. From the data, the calculated RR was 0.73 (95%CI: 0.33 to 1.58). Four patients in the RT-group had a grade 3 oral mucositis. The overall RR for oral mucositis was 0.53 (95%CI: 0.26 to 1.09, TORS: 8 events, RT: 15 events).

Level of evidence of the literature
Transoral robotic surgery (TORS) versus primary radiotherapy (RT)
The level of evidence regarding the outcome measure overall survival (crucial) was downgraded by 3 levels because of the number of included patients (imprecision: only 34 patients were included in each study arm. The number of censored participants was large in both arms, providing limited data to the 5-year overall survival); indirectness (1 level for indirectness: about half of the recruited patients had an N2 carcinoma); publication bias could not be assessed. We did not downgrade for a lack of blinding (risk of bias), because overall survival was considered as a 'hard' outcome measure.

The level of evidence regarding the outcome measure disease-free survival (crucial) was downgraded by 4 levels because of study limitations (1 level for risk of bias: care providers and outcome assessors were not blinded for the treatment allocation); and the number of included patients (2 levels for imprecision: only 34 patients were included in each study arm.

The number of censored participants was large in both arms, providing limited data to the 5-year disease-free survival); indirectness (1 level for indirectness: about half of the recruited patients had an N2 carcinoma); publication bias could not be assessed.

The level of evidence regarding the outcome measure swallowing (crucial) was downgraded by 4 levels because of study limitations (1 level for risk of bias: patients, care providers and outcome assessors were not blinded for the treatment allocation); and the number of included patients (2 levels for imprecision: only 34 patients were included in each study arm); indirectness (1 level for indirectness: about half of the recruited patients had an N2 carcinoma); publication bias could not be assessed.

The level of evidence regarding the outcome measure local control (crucial) was not graded, since there were no studies included that reported this outcome measure.

The level of evidence regarding the outcome quality of life (crucial) was downgraded by 4 levels because of study limitations (1 level for risk of bias: patients, care providers and outcome assessors were not blinded for the treatment allocation); and the number of included patients (2 levels for imprecision: only 34 patients were included in each study arm. Power calculation (using the reported mean difference of 6.8 on de MDADI, SD=12.2, alpha=0.05, beta=0.1, +10% assumed dropout) revealed that 75 participants per arm were needed); indirectness (1 level for indirectness: about half of the recruited patients had an N2 carcinoma); publication bias could not be assessed.

The level of evidence regarding the outcome measure trismus (important) was downgraded by 4 levels because of study limitations (1 level for risk of bias: patients, care providers and outcome assessors were not blinded for the treatment allocation); and the number of included patients (2 levels for imprecision: only 34 patients were included in each study arm); indirectness (1 level for indirectness: about half of the recruited patients had an N2 carcinoma); publication bias could not be assessed.

The level of evidence regarding the outcome measure taste (important) was downgraded by 4 levels because of study limitations (1 level for risk of bias: patients, care providers and outcome assessors were not blinded for the treatment allocation); and the number of included patients (2 levels for imprecision: only 34 patients were included in each study arm); indirectness (1 level for indirectness: about half of the recruited patients had an N2 carcinoma); publication bias could not be assessed.

The level of evidence regarding the outcome measure dryness of the mouth (important) was downgraded by 4 levels because of study limitations (1 level for risk of bias: patients, care providers and outcome assessors were not blinded for the treatment allocation); and the number of included patients (2 levels for imprecision: only 34 patients were included in each study arm); indirectness (1 level for indirectness: about half of the recruited patients had an N2 carcinoma); publication bias could not be assessed.

The level of evidence regarding the outcome measure mucositis (important) was downgraded by 4 levels because of study limitations (1 level for risk of bias: patients, care providers and outcome assessors were not blinded for the treatment allocation); and the number of included patients (2 levels for imprecision: only 34 patients were included in each study arm); indirectness (1 level for indirectness: about half of the recruited patients had an N2 carcinoma); publication bias could not be assessed.

Conclusions

Transoral robotic surgery (TORS, including patients that received adjuvant (chemo)radiotherapy) versus primary radiotherapy (RT, including patients that reveived concurrent chemotherapy and salvage surgery)

Very low GRADE	We are unsure about the differences between TORS and RT in the 5-year overall survival. *Sources: (Nichols, 2019)*

Very low GRADE	We are unsure about the effect of TORS on the disease-free survival when compared to RT. *Sources: (Nichols, 2019)*

Very low GRADE	We are unsure about the effect of TORS on swallowing complaints when compared to RT. *Sources: (Nichols, 2019)*

- GRADE	No studies were included that assessed and reported the local control.

Very low GRADE	We are unsure about the effect of TORS on the quality of life when compared to RT. *Sources: (Nichols, 2019)*

Very low GRADE	We are unsure about the effect of TORS on trismus when compared to RT. *Sources: (Nichols, 2019)*

Very low GRADE	We are unsure about the effect of TORS on taste alterations when compared to RT. *Sources: (Nichols, 2019)*

Very low GRADE	We are unsure about the effect of TORS on the dryness of the mouth when compared to RT. *Sources: (Nichols, 2019)*

Very low GRADE	We are unsure about the effect of TORS on oral mucositis when compared to RT. *Sources: (Nichols, 2019)*

Overwegingen - van bewijs naar aanbeveling

Er werd één gerandomiseerde trial gevonden over Transoral Robotic Surgery (TORS) versus radtiotherapie (RT) die alle uitkomstmaten van interesse rapporteerde, behalve lokale controle (Nichols, 2019). Er was een zeer laag vertrouwen in de gerapporteerde 5-jaars overleving (log-rank p=0,89; HR=0,83 met 95%BHI: 0,21 tot 8,35). Het gerapporteerde

verschil in overleving werd niet als klinisch relevant beschouwd, want hiervoor behoorde de hazard ratio kleiner dan 0,7 te zijn. Het is enigszins onduidelijk hoe de gerapporteerde hazard ratio tot stand is gekomen en welke groep als referentiegroep is gekozen in de calculatie hiervan. Ook werd er geen reden aangedragen waarom er een groot aantal deelnemers uit beide interventiegroepen gecensureerd werden in de analyse. Vermoedelijk verlieten deelnemers het onderzoek (de mediane follow-up was 27 maanden), maar het gerapporteerde verlies in de follow-up in de studie-doorstroom was slechts 2 personen in de primaire radiotherapie groep. Eén patiënt in de studie overleed ten gevolge van een Transoral Robotic Surgery (TORS)-ingreep gerelateerde bloeding. Er was verder een zeer laag vertrouwen in de gerapporteerde ziektevrije overleving (HR=1,07 met 95%BHI: 0,28 tot 4,01), slikklachten (RR=1,11 met 95%BHI: 0.90 tot 1.37; Eén persoon in de studie (radiotherapie-groep) had een percutane voedingssonde) en kwaliteit van leven (MDADI totale score: 80,1 (TORS) versus 86,9 (RT), p=0,042; MDADI composietscore: 80,2 (TORS) versus 86,7 (RT), p=0,049). Er werden significante verschillen in (slik-gerelateerde) kwaliteit van leven met het MDADI-instrument gemeten, maar deze konden niet worden gezien als een klinisch relevant verschil (dat wil zeggen >10 punten verschil). Kwaliteit van leven werd tevens met meerdere vragenlijsten beoordeeld. Zie Tabel 12.1 in de resultaten onder het tabblad 'Onderbouwing' voor de door Nichols (2019) gerapporteerde scores. Het zeer lage vertrouwen werd veroorzaakt door het geringe aantal deelnemers in de studie en het gebrek aan blindering van patiënten, zorgverleners en uitkomstbeoordelaars bij 'zachte' uitkomstmaten. Er werden geen resultaten gevonden voor lokale controle. Ook in de gevonden resultaten over de belangrijke uitkomstmaten is er op dit moment een zeer laag vertrouwen door het geringe aantal deelnemers in de studie en het gebrek aan blindering van patiënten, zorgverleners en uitkomstbeoordelaars.

Nichols (2019) berekende het benodigde aantal deelnemers aan de hand van het klinisch relevante verschil op de MDADI (gemiddeld verschil: 10 punten; gepoolde SD: 12; alfa=0,05, bèta=0,1, 10% verwachte drop-out). Het aantal deelnemers in de studie voldeed aan deze berekening, maar er werd geen klinisch relevant verschil gevonden. Aan de hand van het daadwerkelijk gevonden verschil kon worden berekend dat de behaalde 'power' 0,57 was bij een gemiddeld verschil van 6,8 punten, een gepoolde SD van 12,3, een alfa van 0,05, een bèta van 0,9, en met n=30 per arm (er waren 30 (TORS) en 27 (RT) ingevulde vragenlijsten op het 1-jaars tijdpunt). Omdat er een zeer beperkt aantal patiënten aan deze studie deelnamen is de imprecisie van de effectschatters groot. In de toekomst zullen resultaten van andere gerandomiseerde studies meer informatie toevoegen over het eventuele (on)gunstige effect van transorale robotische chirurgie ten opzichte van primaire radiotherapie. Via het zoekportaal in het International Clinical Trials Registry Platform (http://apps.who.int/trialsearch/Default.aspx) van de World Health Organization werd er naar geregistreerde gerandomiseerde studies gezocht die transorale robotische chirurgie vergeleken met een vorm van primaire radiotherapie. Er werden vijf gerandomiseerde trials gevonden (zie Tabel 12.2), waarvan er 1 reeds werd beëindigd, 2 actief zijn maar niet rekruteren en 2 op dit moment rekruteren. Naar verwachting zullen de laatste metingen van patiënten in de trials plaats vinden tussen 2021 en 2029.

Tabel 12.2 Geïdentificeerde trials naar transorale robotische chirurgie versus radiotherapie voor orofarynx carcinomen; uit het International Clinical Trials Registy Platform van de World Health Organization

Randomized trial	Trial ID	Status	Intervention A	Intervention B	Inclusion TNM	Primary completion date (last participant	Study completion date (last measurement of interventio

						recruitment)	n / treatment)
EORTC-1420-HNCG-ROG (“Best of” trial)	NCT02984410 (Clinicaltrials.gov)	Recruiting	Any transoral surgery approach (e.g. transoral laser micro-surgery, conventional transoral surgery, transoral robotic surgery)	Intensity modulated radiotherapy with simultaneous integrated boost		June 2021	January 2027
QoLATI	NCT04124198 (Clinicaltrials.gov)	Recruiting	Transoral robotic surgery + neck dissection	Intensity-modulated radiation therapy	cT1-2 cN0-1 Distant metastases will be excluded	January 2024	January 2029
ECOG 3311	NCT01898494 (Clinicaltrials.gov)	Active, not recruiting	Transoral surgery	Transoral surgery + low-dose intensity-modulated radiation therapy Transoral surgery + standard-dose intensity-modulated radiation therapy Transoral surgery + standard-dose intensity-modulated radiation therapy + chemotherapy	Stage III Stage IVa Stage IVb No evidence of distant metastases	February 2020	February 2023
ORATOR	NCT01590355 (Clinicaltrials.gov)	Active, not recruiting	Transoral robotic surgery + neck dissection	Radiotherapy with or without chemotherapy	T1-2 N0-2 Without extranodal extension	June 2021	June 2021
NRG/RTOG 1221	NCT01953952 (Clinicaltrials.gov)	Withdrawn (Slow accrual)	-	-	-	-	-

In 2009 werd TORS goedgekeurd voor de behandeling T1-2 orofarynxcarcinomen door de Amerikaanse Food and Drug Administration (FDA) op basis van korte termijn uitkomsten (30-dagen). De volgende studies hebben destijds tot de goedkeuring geleid en zijn samengevat in Tabel 12.3. (Weinstein, 2007; Weinstein, 2007). Zowel TORS als radiotherapie hebben een curatief doel. Beide behandelingen hebben echter een verschillend patroon van complicaties en toxiciteit door de aard van de interventie. De voorkeur van een patiënt voor één van beide interventies kan hiervan afhangen. Het is daarom belangrijk om de voor- en nadelen van beide interventies met de patiënt te bespreken.

Tabel 12.3 Uitkomsten van studies die tot de FDA goedkeuring hebben geleid

Author (year)	Intervention	Sample size	Patient characteristics	Outcomes
Weinstein (2007)	Radical tonsillectomy	n=27 (1 patient lost to follow-up)	Sex: 25 males / 2 females T-stage, n: T1: 5 T2: 16 T3: 6 N-stage, n: N0: 4 N1: 13 N2: 10 N3: 0 Differentiation, n: Well: 2 Moderate to well: 2 Moderate: 11 Moderate to poor: 6 Poor: 6 Karnofsky score, n: 0-60: 0 70: 1 80: 3 90: 15 100: 8 Charlson Comorbidity Index, n: 0: 11 1: 12 2: 2 3: 3	*During TORS procedure:* Mean blood loss, ml (range): 189 (0-500) *30-day postoperative complications:* Mucosal bleeding, n: 1/27 (4%) Tracheotomy (for exacerbation of sleep apnea), n: 1/27 (4%) Moderate trismus, n: 2/27 (7%) Hypernasality, n: 1/27 (4%) Delirium tremens, n: 1/27 (4%) Mortality, n: 0/27 (0%) *Follow-up:* Local or regional recurrences, n: 0/26 (0%) Intubation period: 1 patient had a tracheotomy during the TORS procedure 20 patients were extubated at the end of the TORS procedure 6 patients remained intubated for 2.7 days postoperatively (range: 2-3 days) Swallowing without the use of gastrostomy, n: 26/27 (96%) Adjuvant therapy, n: Postoperative radiation without chemotherapy: 9 Postoperative irradiation with chemotherapy: 15 Postoperative chemotherapy: 1
Weinstein (2007)	Supraglottic partial laryngectomy	n=3	Sex: 2 males / 1 female	Surgical time, hours:minutes:seconds: Patient 1: 2:58:18 Patient 2: 1:35:01

			Median age, years: 62.3 TNM-stage, n: T2N0M0: 2 T3N0M0: 1	Patient 3: 1:32:48 <u>Hospitalization, days:</u> Patient 1: 3 Patient 2: 8 Patient 3: 5 <u>Return to swallowing, week:</u> Patient 1: 6 Patient 2: 5 Patient 3: 5 <u>Mean blood loss, ml (range):</u> 200 (100-400) <u>Complications:</u> No complications <u>Conversions:</u> No conversions <u>Adjuvant therapy:</u> The patient with the T3N0M0 carcinoma received chemoradiation after positive lymph nodes were found from neck dissection

TORS heeft als voordeel dat het één enkele ingreep betreft en, wanneer er geen complicaties optreden, de opnameduur één nacht is. Wanneer bij histopathologisch onderzoek van het resectiepreparaat na TORS geen indicaties voor adjuvante radiotherapie bestaat kan radiotherapie achter de hand worden gehouden voor het geval er recidief of tweede primaire tumor optreedt. Echter, een nadeel van TORS is dat de patiënt onder narcose moet om de operatie te ondergaan. Er kunnen bloedingen, pijn en ontstekingen ontstaan. Na de ingreep kan tevens blijken dat TORS niet afdoende is geweest en zal een nabehandeling met (chemo)radiatie noodzakelijk zijn. Wanneer er gekozen wordt voor een behandeling met TORS bestaat er een kans dat adjuvante behandeling noodzakelijk blijkt. In de RCT van Nichols (2019), waarin de helft (51%) van de deelnemers N2 status hadden, kregen 70,6% van de deelnemers die TORS ondergingen óók postoperatieve radiotherapie (n=16; 47,1%) of chemoradiotherapie (n=8; 23,5%). De kans op adjuvante chemoradiotherapie zou hoogstwaarschijnlijk lager zijn wanneer er geen patiënten met N2 status waren gerekruteerd. Voor radiotherapie hoeven patiënten niet onder narcose. Door de aard van deze interventie levert radiotherapie géén chirurgie-gerelateerde complicaties op, zoals bloedingen en ontstekingen. Bij het verstrekken van radiotherapie is de duur van de behandeling langer dan de chirurgische interventie. Een radiotherapeutische behandeling kan tot zeven weken duren. Radiotherapie kan daarnaast acute toxiciteit veroorzaken met slikklachten, smaakverlies en een droge mond als gevolg. Ook kunnen lange termijn bijwerkingen optreden zoals, eveneens, slikklachten en een droge mond, maar ook hypothyroïdie, radiatiecariës, radionecrose en/of versnelling van de atherosclerose. Als gevolg van de radiotherapie kunnen met zeer geringe kans op langere termijn secundaire tumoren ontstaan in het hoofd-halsgebied. Van de patiënten die in de RCT van Nichols (2019) radiotherapie ondergingen, in plaats van TORS, ontving 71,9% (n=23) óók concomitante chemotherapie.

De kosteneffectiviteit van deze therapieën zouden een verdere afweging kunnen zijn. De Almeida (2015) beschreef dat TORS kosteneffectief is bij patiënten met vroege T-stadium tumoren ongeacht het N-stadium, maar dat de kosteneffectiviteit afneemt naarmate er

vaker adjuvante therapie na TORS gegeven moet worden. De auteurs stellen dat een goede patiëntenselectie hierbij belangrijk is om de waarschijnlijkheid op een adjuvante behandeling te verminderen. Rudmik (2015) rapporteert dat intensiteitsgemoduleerde radiotherapie een grotere waarschijnlijkheid heeft om kosteneffectief te zijn dan TORS bij patiënten met T1-2N0M0 tumoren bij afkapwaarden van de bereidheid tot betalen tussen $50.000 en $150.000 per quality-adjusted life-year (QALY). De auteurs dragen aan dat hoog-volume centra wellicht de waarde van zorg (zgn. value of care) van TORS zou kunnen verhogen. Rodin (2016) beschrijft dat de kosteneffectiviteit van zowel TORS als RT afhankelijk waren van specifieke klinische situaties bij niet eerder behandelde patiënten met cT1-2cN0-1 tumoren. De auteurs rapporteren door middel van een probabilistische analyse dat RT een grotere waarschijnlijkheid heeft om kosteneffectief te zijn bij afkapwaarden van de bereidheid tot betalen tussen $50.000 en $150.000 per QALY. Sher (2016) beschreef dat RT een kosteneffectieve behandeling met hoge waarde was voor de niet-rokende, 65 jaar oude man met een HPV-positieve T1-2N2a-b tumor. De auteurs gaven aan dat TORS wellicht een kosteneffectieve interventie zou kunnen worden wanneer het tot een betekenisvolle relatieve verbetering van locoregionale controle zou leiden ten opzichte van primaire chemoradiotherapie, maar dat er nog geen literatuur bekend was die dit aan zou tonen. Kosteneffectiviteitsanalyses zijn afhankelijk van de gebruikte parameters in het model en de eventuele aannames in de analyses. De hier besproken kosteneffectiviteitsanalyses zijn daarom wellicht niet direct één-op-één over te nemen voor de Nederlandse situatie.

In Nederland wordt TORS uitgevoerd in een beperkt aantal geselecteeerde centra voor beperkte indicaties. Er zijn geen algemeen geaccepteerde indicaties maar vaak uitgeoefende selectiecriteria zijn de volgende: de tumor is goed bereikbaar, een excisie met chirurgische marges (1cm) is haalbaar en er is geen of één kleine lymfekliermetastase in de hals. De overweging is om adjuvante RT na TORS te voorkomen.

Op dit moment is radiotherapie de standaardbehandeling in Nederland. Het is van belang dat de patiënt hiervan op de hoogte wordt gebracht, maar ook dat er eventueel een alternatieve chirurgische optie mogelijk is middels TORS. Door de voor- en nadelen van beide interventies te bespreken in combinatie met de professionele opinie van de medisch-specialist zal gezamenlijk (middels gedeelde besluitvorming) moeten worden afgewogen welke behandeling zal worden ingezet.

De werkgroep is echter van mening dat radiotherapie op dit moment de standaard zorg blijft, omdat er onvoldoende bewijs werd gevonden wat met hogere zekerheid het afwijken van de huidige standaardbehandeling ten faveure van TORS in voldoende mate ondersteunt. Er zijn een aantal lopende gerandomiseerde studies geïdentificeerd (Tabel 12.2). Resultaten van die studies zullen meer wetenschappelijke data toevoegen zodat er in de toekomst mogelijk met hogere zekerheid aanbevelingen kunnen worden gemaakt met betrekking tot de (on)gunstige effecten van beide interventies.

Door een sterke voorkeur van de patiënt voor een chirurgische behandeling zou TORS als alternatieve interventie eventueel ingezet kunnen worden wanneer TORS in dat geval óók door de medisch specialist als adequaat voor de betreffende situatie wordt beschouwd. De chirurgische optie vervalt vanzelfsprekend indien er contra-indicaties voor TORS aanwezig zijn. Om de chirurgische interventie (dat wil zeggen TORS) te kunnen uitvoeren moet de betreffende chirurg de benodigde certificering hebben. Daarnaast moet de chirurg ook ervaring met deze operatie hebben en de operatie op regelmatige basis uitvoeren om TORS te kunnen aanbieden.

Aanbevelingen

Aanbeveling-1

Rationale van de aanbeveling: weging van argumenten voor en tegen de interventies

De werkgroep is van mening dat patiënten door de verschillende complicatie- en toxiciteitspatronen die ontstaan wegens de aard van beide interventies een sterke voorkeur voor één van beide interventies kan hebben. De werkgroep acht het daarom belangrijk dat de voor- en nadelen van beide interventies besproken moeten worden, maar ook dat aangegeven wordt dat radiotherapie vooralsnog de huidige standaardzorg is.

Ondersteun de patiënt bij het maken van een behandelkeuze waarbij de individuele patiëntkarakteristieken dienen te worden afgewogen.

En bespreek de volgende belangrijke voor- en nadelen van de interventies indien van toepassing op de patiënt:

- **Behandelduur:**
  De duur van de behandelingen zonder complicaties bij TORS is één ingreep met een opnameduur van één tot enkele dagen. Voor RT is de behandelduur tot zeven weken.
- **Procedures:**
  TORS wordt onder narcose verricht, RT niet. RT zal volgens een behandelschema verlopen.
- **Korte en lange termijn complicaties en toxiciteit:**
  Bij TORS kunnen bloedingen, pijnklachten en ontstekingen optreden. Bij RT kan smaakverlies, een droge mond en slikklachten optreden. Ook kunnen hypothyroïdie, versnelling van artherosclerose en, met een zeer gering risico, secundaire tumoren ontstaan ten gevolge van RT. De kans op radiatiecariës en radionecrose neemt toe naar mate de tumor dicht bij de mandibula ligt of de manibula invalideert.
- **Kans op adjuvante behandeling:**
  Voor TORS bestaat er een aanzienlijke kans op een adjuvante behandeling met (chemo)radiatie.

Aanbeveling-2

Rationale van de aanbeveling: weging van argumenten voor en tegen de interventies

De werkgroep is van mening dat radiotherapie de standaardbehandeling van patiënten met T1-2N0-1 orofarynxcarcinomen blijft. Op dit moment is de huidige behandeling radiotherapie en geeft de lage zekerheid in het gevonden bewijs vooralsnog geen richting om TORS boven radiotherapie te verkiezen, maar behandel patiënten zo veel mogelijk met één behandelingsmodaliteit. Bij patiënten met N1 ziekte die behandeld worden met TORS zal er een halsklierdissectie verricht moeten worden. De resultaten van de lopende gerandomiseerde studies dienen afgewacht te worden.

Deze aanbeveling is daarom gebaseerd op het ontbreken van een hoge(re) zekerheid in alle eindpunten die in de PICO gedefinieerd zijn. Er is maar één RCT beschikbaar. Vanuit deze studie zijn er met lage zekerheid aanwijzingen dat beide behandelingen (TORS en RT) mogelijk een vergelijkbare 5-jaars overleving hebben. Er is geen data over de impact van beide modaliteiten wat betreft de lokale controle in deze studie. Tevens is het effect van TORS ten opzichte van RT op alle toxiciteit eindpunten onduidelijk vanuit deze studie.

Geef radiotherapie bij patiënten met primaire T1-2N0-1 orofarynxcarcinomen, tenzij er patiëntvoorkeuren voor een chirurgische behandeling zijn.

Literatuur

De Almeida JR, Moskowitz AJ, Miles BA, Goldstein DP, Teng MS, Sikora AG, Gupta V, Posner M, Genden EM. Cost-effectiveness of transoral robotic surgery versus (chemo)radiotherapy for early T classification oropharyngeal carcinoma: A cost-utility analysis. Head Neck. 2016 Apr;38(4):589-600. doi: 10.1002/hed.23930. Epub 2015 Jun 30. PMID: 25488048.

Nichols AC, Yoo J, Hammond JA, Fung K, Winquist E, Read N, Venkatesan V, MacNeil SD, Ernst DS, Kuruvilla S, Chen J, Corsten M, Odell M, Eapen L, Theurer J, Doyle PC, Wehrli B, Kwan K, Palma DA. Early-stage squamous cell carcinoma of the oropharynx: radiotherapy versus. trans-oral robotic surgery (ORATOR)—study protocol for a randomized phase II trial. BMC Cancer. 2013 Mar 20;13:133. doi: 10.1186/1471-2407-13-133. PubMed PMID: 23514246; PubMed Central PMCID: PMC3621077.

Nichols AC, Theurer J, Prisman E, Read N, Berthelet E, Tran E, Fung K, de Almeida JR, Bayley A, Goldstein DP, Hier M, Sultanem K, Richardson K, Mlynarek A, Krishnan S, Le H, Yoo J, MacNeil SD, Winquist E, Hammond JA, Venkatesan V, Kuruvilla S, Warner A, Mitchell S, Chen J, Corsten M, Johnson-Obaseki S, Eapen L, Odell M, Parker C, Wehrli B, Kwan K, Palma DA. Radiotherapy versus transoral robotic surgery and neck dissection for oropharyngeal squamous cell carcinoma (ORATOR): an open-label, phase 2, randomised trial. Lancet Oncol. 2019 Oct;20(10):1349-1359. doi: 10.1016/S1470-2045(19)30410-3. Epub 2019 Aug 12.

Erratum in: Lancet Oncol. 2019 Dec;20(12):e663. PubMed PMID: 31416685.

NVMO-commissie ter Beoordeling van Oncologische Middelen (BOM). PASKWIL criteria 2018.

Rodin D, Caulley L, Burger E, Kim J, Johnson-Obaseki S, Palma D, Louie AV, Hansen A, O'Sullivan B. Cost-Effectiveness Analysis of Radiation Therapy Versus Transoral Robotic Surgery for Oropharyngeal Squamous Cell Carcinoma. Int J Radiat Oncol Biol Phys. 2017 Mar 15;97(4):709-717. doi: 10.1016/j.ijrobp.2016.11.029. Epub 2016 Nov 27. PMID: 28244405.

Rudmik L, An W, Livingstone D, Matthews W, Seikaly H, Scrimger R, Marshall D. Making a case for high-volume robotic surgery centers: A cost-effectiveness analysis of transoral robotic surgery. J Surg Oncol. 2015 Aug;112(2):155-63. doi: 10.1002/jso.23974. Epub 2015 Jul 14. PMID: 26171771.

Sher DJ, Fidler MJ, Tishler RB, Stenson K, al-Khudari S. Cost-Effectiveness Analysis of Chemoradiation Therapy Versus Transoral Robotic Surgery for Human Papillomavirus-Associated, Clinical N2 Oropharyngeal Cancer. Int J Radiat Oncol Biol Phys. 2016 Mar 1;94(3):512-22. doi: 10.1016/j.ijrobp.2015.11.006. Epub 2015 Nov 10. PMID: 26867880.

Weinstein GS, O'Malley BW Jr, Snyder W, Sherman E, Quon H. Transoral robotic surgery: radical tonsillectomy. Arch Otolaryngol Head Neck Surg. 2007 Dec;133(12):1220-6. doi: 10.1001/archotol.133.12.1220. PMID: 18086963.

Weinstein GS, O'Malley BW Jr, Snyder W, Hockstein NG. Transoral robotic surgery: supraglottic partial laryngectomy. Ann Otol Rhinol Laryngol. 2007 Jan;116(1):19-23. doi: 10.1177/000348940711600104. PMID: 17305273.

Geldigheid en Onderhoud

Module[1]	Regiehouder(s)[2]	Jaar van autorisatie	Eerstvolgende beoordeling actualiteit richtlijn[3]	Frequentie van beoordeling op actualiteit[4]	Wie houdt er toezicht op actualiteit[5]	Relevante factoren voor wijzigingen in aanbeveling[6]

[1] Naam van de module
[2] Regiehouder van de module (deze kan verschillen per module en kan ook verdeeld zijn over meerdere regiehouders)
[3] Maximaal na vijf jaar
[4] (half)Jaarlijks, eens in twee jaar, eens in vijf jaar
[5] regievoerende vereniging, gedeelde regievoerende verenigingen, of (multidisciplinaire) werkgroep die in stand blijft
[6] Lopend onderzoek, wijzigingen in vergoeding/organisatie, beschikbaarheid nieuwe middelen

Bijlagen bij module 12.1

Kennislacunes
What are the (un)beneficial effects of TORS (with or without microsurgery or adjuvant radiotherapy) on overall survival, disease-free survival, local control, swallowing, trismus, taste, dryness of the mouth, mucositis, and quality of life in patients with a T1-2N0-1 oropharyngeal carcinoma, when compared to primary radiotherapy.

Indicatoren
Geen.

Implementatieplan

Aanbeveling	Tijdspad voor implementatie: < 1 jaar, 1 tot 3 jaar of 3 tot 5 jaar	Verwacht effect op kosten	Randvoorwaarden voor implementatie (binnen aangegeven tijdspad)	Mogelijke barrières voor implementatie[1]	Te ondernemen acties voor implementatie[2]	Verantwoordelijken voor acties[3]	Overige opmerkingen
Ondersteun de patiënt bij het maken van een behandelkeuze waarbij de individuele patiëntkarakteristieken dienen te worden afgewogen. En bespreek de volgende belangrijke voor- en nadelen van de interventies indien van toepassing op de patiënt: • Behandelduur: *De duur van de behandeling en zonder complicaties bij TORS is één ingreep met een opnameduur van 1 nacht. Voor RT is de behandelduur tot zeven weken.* • Procedures: *TORS wordt onder narcose verricht, RT*	< 1 jaar	Geen	Geen	Geen	Geen		

niet. RT zal volgens een behandelschema verlopen. • Korte en lange termijn complicaties en toxiciteit: *Bij TORS kunnen bloedingen, pijnklachten en ontstekingen optreden. Bij RT kan smaakverlies, een droge mond en slikklachten optreden. Ook kunnen hypothyroïdie, versnelling van artherosclerose en secundaire tumoren ontstaan ten gevolge van RT.* • Kans op adjuvante behandeling: *Voor TORS bestaat er een aanzienlijke kans op een adjuvante behandeling met (chemo)radiatie.*							
Geef radiotherapie bij patiënten met primaire T1-2N0-2 orofarynxcarcinomen, tenzij er sterke patiëntvoorkeuren voor een chirurgische behandeling zijn.	< 1 jaar	Geen	Geen	Geen	Geen		

[1] Barrières kunnen zich bevinden op het niveau van de professional, op het niveau van de organisatie (het ziekenhuis) of op het niveau van het systeem (buiten het ziekenhuis). Denk bijvoorbeeld aan onenigheid in het land met betrekking tot de aanbeveling, onvoldoende motivatie of kennis bij de specialist, onvoldoende

faciliteiten of personeel, nodige concentratie van zorg, kosten, slechte samenwerking tussen disciplines, nodige taakherschikking, et cetera.
[2] Denk aan acties die noodzakelijk zijn voor implementatie, maar ook acties die mogelijk zijn om de implementatie te bevorderen. Denk bijvoorbeeld aan controleren aanbeveling tijdens kwaliteitsvisitatie, publicatie van de richtlijn, ontwikkelen van implementatietools, informeren van ziekenhuisbestuurders, regelen van goede vergoeding voor een bepaald type behandeling, maken van samenwerkingsafspraken.
[3] Wie de verantwoordelijkheden draagt voor implementatie van de aanbevelingen, zal tevens afhankelijk zijn van het niveau waarop zich barrières bevinden. Barrières op het niveau van de professional zullen vaak opgelost moeten worden door de beroepsvereniging. Barrières op het niveau van de organisatie zullen vaak onder verantwoordelijkheid van de ziekenhuisbestuurders vallen. Bij het oplossen van barrières op het niveau van het systeem zijn ook andere partijen, zoals de NZA en zorgverzekeraars, van belang.

Evidence tables

Study reference (first author, publication year)	Describe method of randomisation[1]	Bias due to inadequate concealment of allocation?[2] (unlikely/likely/unclear)	Bias due to inadequate blinding of participants to treatment allocation?[3] (unlikely/likely/unclear)	Bias due to inadequate blinding of care providers to treatment allocation?[3] (unlikely/likely/unclear)	Bias due to inadequate blinding of outcome assessors to treatment allocation?[3] (unlikely/likely/unclear)	Bias due to selective outcome reporting on basis of the results?[4] (unlikely/likely/unclear)	Bias due to loss to follow-up?[5] (unlikely/likely/unclear)	Bias due to violation of intention to treat analysis?[6] (unlikely/likely/unclear)
Nichols 2019	Random 1:1 (block size 4) assignment using a computer-generated randomisation list. The list was prepared by a statistician, who was not involved in the study. Groups were stratified by p16-status (hpv surrogate marker)	Unlikely Reason: Assignment was done by a trial coordinator who was not involved in clinical management. Only the trial coordinator had access to the locked and concealed list of allocations. Allocation of treatment group was communicated by mail. The block-size of 4 might be too small for complete concealment. However, only the statistician knew	Unlikely (for overall survival) Reason: Overall survival was seen as a hard outcome measure not affected by blinding. Likely (for progression-free survival, complications, toxicity, swallowing function, quality of life) Reason: Authors state that patients were not blinded.	Unlikely (for overall survival) Reason: Overall survival was seen as a hard outcome measure not affected by blinding. Likely (for progression-free survival, complications, toxicity, swallowing function, quality of life) Reason: Authors state that physicians were not blinded.	Unlikely (for overall survival) Reason: Overall survival was seen as a hard outcome measure not affected by blinding. Likely (for progression-free survival, complications, toxicity) Unclear (for quality of life) Reason: It is unclear who was the assessor of the quality-of-life questionnaires, although it might presumably be the treating physician.	Unlikely Reason: All outcomes of interest were reported in the protocol and the published study report.	Unlikely Reason: Two persons (2/34, 5.9%) withdrew from the study after randomization in the radiotherapy group.	Unlikely Reason: Authors state that an ITT analysis was performed. There were no indications that participants deviated from the protocol. However, there is some missing data other than loss to follow-up and no reason was provided. Data was not imputed, and cases were thus excluded from analysis (e.g., 27/34 (RT) and 30/34 (TORS) datapoints were available for the

		the exact block size.						primary outcome at 1 year); There is more missing data than withdrawal (figure 1 in the study report) and mortality (figure 3) account for.

1. **Randomisation: generation of allocation sequences have to be unpredictable, for example computer generated random-numbers or drawing lots or envelopes. Examples of inadequate procedures are generation of allocation sequences by alternation, according to case record number, date of birth or date of admission.**
2. **Allocation concealment: refers to the protection (blinding) of the randomisation process. Concealment of allocation sequences is adequate if patients and enrolling investigators cannot foresee assignment, for example central randomisation (performed at a site remote from trial location) or sequentially numbered, sealed, opaque envelopes. Inadequate procedures are all procedures based on inadequate randomisation procedures or open allocation schedules.**
3. **Blinding: neither the patient nor the care provider (attending physician) knows which patient is getting the special treatment. Blinding is sometimes impossible, for example when comparing surgical with non-surgical treatments. The outcome assessor records the study results. Blinding of those assessing outcomes prevents that the knowledge of patient assignement influences the proces of outcome assessment (detection or information bias). If a study has hard (objective) outcome measures, like death, blinding of outcome assessment is not necessary. If a study has "soft" (subjective) outcome measures, like the assessment of an X-ray, blinding of outcome assessment is necessary.**
4. **Results of all predefined outcome measures should be reported; if the protocol is available, then outcomes in the protocol and published report can be compared; if not, then outcomes listed in the methods section of an article can be compared with those whose results are reported.**
5. **If the percentage of patients lost to follow-up is large, or differs between treatment groups, or the reasons for loss to follow-up differ between treatment groups, bias is likely. If the number of patients lost to follow-up, or the reasons why, are not reported, the risk of bias is unclear.**
6. **Participants included in the analysis are exactly those who were randomized into the trial. If the numbers randomized into each intervention group are not clearly reported, the risk of bias is unclear; an ITT analysis implies that (a) participants are kept in the intervention groups to which they were randomized, regardless of the intervention they actually received, (b) outcome data are measured on all participants, and (c) all randomized participants are included in the analysis.**

Evidence table for intervention studies (randomized controlled trials and non-randomized *observational* studies (cohort studies, case-control studies, case series))[1]

This table is also suitable for diagnostic studies (screening studies) that compare the effectiveness of two or more tests. This only applies if the test is included as part of a test-and-treat strategy - otherwise the evidence table for studies of diagnostic test accuracy should be used.

Study reference	Study character-istics	Patient characteristics [2]	Intervention (I)	Comparison / control (C) [3]	Follow-up	Outcome measures and effect size [4]	Comments
Nichols 2019	Type of study: RCT (multicentre, international) Setting and country: Hospital, Canada and Australia Funding and conflicts of interest: Canadian cancer society research institute grant, Ontario institute for carcer reseach clinicion-scientist grant, Wolfe surgical research professorship in the biology of head and neck cancers grant. One author received honoraria and educational grants from Eisai and Genzyme-	Inclusion criteria: From the Protocol (Nichols 2013, BMJ): >18 years, providing informed consent, ECOG status 0-2, histologically confirmed SCC, primary site is the oropharynx, T1-2 with likely negative resection margins at surgery, N0-2 without extranodal extension on pre-randomization imaging, CBC/differential obtained within 4 weeks prior of randomization (with adequate bone marrow, hepatic, and renal function), patient was assessed at a head and neck multidisciplinary clinic and presented at a multidisciplinary tumor board prior to randomization.	Describe intervention (treatment/procedure/test): TORS + Neck dissection: A surgical robot (Da Vinci) was used to surgically excise the primary oropharyngeal tumor with the spatula cautery (resection margin: 1cm). Neck dissection was performed selectively at the time of surgery or within 2 weeks of surgery (at the discretion of the surgeon). Surgeons had to have at least 10 TORS operations performed prior to enrolling patients to the trial. Surgical specimen was sent for frozen section analysis. The resection proceeded until negative margins were obtained.	Describe control (treatment/procedure/test): Radiotherapy (intensity modulated): Radiotherapy was given. Gross tumor and nodes: 70Gy / 35 fractions / 7 weeks Region of positive margins/ estranodal extension: N/A High risk nodal area: 63Gy / 35 fractions / 7 weeks Low risk nodal area: 56Gy / 35 fractions / 7 weeks Accelerated or hyperfractionated regimens (i.e., same dose at 6 weeks) could have been used at the radiation oncologist's discretion. Concurrent chemotherapy was given when T1-2N1-2 and omitted when T1-2N0. In the additional files the authors state that cisplatin 100mg/m^2 was delivered	Length of follow-up: Every 3 months in the first 2 years. Every 6 months in year 3-5. (Median follow-up was 27 months, IQR: 20-48) Physical examination: all visits Adverse events monitoring: all visits. QoL measurement: every 6 months Chest x-ray: every 6 months (except at T=year 1, where a CT was taken) In the RT group: additional scan at 8-12 weeks after radiotherapy to assess treatment response / need for salvage surgery. Loss-to-follow-up: Intervention (TORS+ND): 0	Outcome measures and effect size (include 95%CI and p-value if available): **Overall survival:** 5-year overall survival: TORS: 2 at risk (27 censored) RT: 2 at risk (28 censored) RT/TORS: HR = 0.83 (95%CI: 0.21-8.35), log-rank p-value = 0.89 Unclear why patients were censored. One patient in the TORS group died of bleeding. **Disease-free survival:** 5-year progression-free survival: Progression-free survival was defined as time from randomization to either death or recurrence of disease (whichever occurred first in time) TORS: 2 at risk (27 censored) RT: 2 at risk (28 censored)	The authors describe that there was an ITT analysis performed. Trial flow (figure 1) shows that n=68 patients were analysed. However, some tables and figures show other numbers analysed (e.g., table 2 where clinically meaningful decline was analysed in 26 or 27 patients per group). Authors chose not to impute data.

	Sanofi (declared as unrelated to the RCT). All other authors declare no CoI.	Exclusion criteria: From the Protocol (Nichols 2013, BMJ): Serious medical comorbidities, contraindications to radio therapy / chemotherapy / surgery, prior history of head and neck cancer within 5 years, prior head and neck radiation, metastatic disease, inability to attend full-course radiotherapy or follow-up visits, prior invasive malignant disease (unles 5 year disease-free) with the exception of non-melanoma skin cancer, unable or unwilling to complete QoL questionnaires, pregnant or lactating. N total at baseline: Intervention: 34 Control: 34 Important prognostic factors[2]: *Median age (IQR):* *TORS: 58.2 (52.6-64.5)* *C: 60 (53.2-65.2)*	Adjuvant radiotherapy (intensity modulated radiotherapy) was given. Radiotherapy could be omitted when there were no extranodal extensions, positive margins, pT3-4, nodal disease, or lympovascular invasion. RT doses and fractionations: Gross tumor and nodes: N/A Region of positive margins/ estranodal extension: 64 Gy / 30 fraction / 6 weeks High risk nodal area: 60Gy / 30 fractions / 6 weeks Low risk nodal area: 54Gy / 30 fractions / 6 weeks Adjuvant chemotherapy (cisplatin, 100mg/m^2 was delivered every 3 weeks in a 3-week cycle) was given concurrent with radiotherapy when there were positive margins or extra-capsular extension. Doses and/or schedule	every 3 weeks in a 3-week cycle. Doses and/or schedule for patients deemed unfit for cisplatin could be modified, or cetuximab or weekly carboplatin could be administered. Treatment response was evaluated 8 to 12 weeks after completion of radiotherapy by CT or PET-CT. For CT, patients with residual nodes >1cm were recommended for salvage surgery. For PET-CT, salvage surgery was recommended for FDG-avid nodes >1cm. Patients having a relapse or progressive disease following RT, surgical salvage was offered when feasible. *** Out of 34 persons, 32 (94%) received radiotherapy (2 lost to follow-up). Twenty-three (68%) persons received chemotherapy (cisplatin: 19/23 (83%), carboplatin: 3/23 (13%), cetuximab: 1/23 (4%))	Reasons (describe) Control (RT): 2 N=2 (5.9%) Reasons (describe): not provided. Incomplete outcome data: For primary endpoint at T0: Intervention (TORS+ND): 3 N=3 (8.8%) Reasons (describe): no reasons provided. Control (RT): 2 N=2 (5.9%) Reasons (describe) For primary endpoint at T1: Intervention (TORS+ND): 4 N=4 (11.8%) Reasons (describe): No reasons described. However, from survival analysis: 2 persons died. O Control (RT): 7 N=7 (20.6%)	RT/TORS: HR = 1.07 (95%CI: 0.28-4.01), log-rank p-value = 0.63 Unclear why patients were censored. **Swallowing:** Grade 1-2 dysphagia, n (%): TORS: 21/34 (61.8%) RT: 21/34 (61.8%) RR (TORS/RT) = 1.00 (95%CI: 0.69-1.45) Grade 3 dysphagia, n (%): TORS: 9/34 (26.5%) RT: 6/34 (17.6%) RR (TORS/RT) = 1.50 (95%CI: 0.60-3.75) All grade dysphagia, n (%): TORS: 30/34 (88.2%) RT: 27/34 (79.4%) RR (TORS/RT) = 1.11 (95%CI: 0.90-1.37) Use of percutaneous feeding tube at 1 year, n (%): TORS: 0/34 (0%) RT: 1/34 (3%) **Trismus:** Grade 1-2 trismus, n (%): TORS: 8/34 (23.5%) RT: 1/34 (2.9%)	

		Sex (M/F): *TORS: 28M/6F* *RT: 31M/3F* *Smoking history, n (%):* *TORS: 21 (62%)* *RT: 28 (82%)* *Over 21 drinks per week, n (%):* *TORS: 6/22 (27%)* *RT: 1/18 (6%)* *Tumor location:* *TORS: tonsil/ tonsilar foss (n=24), base of tongue (n=10)* *RT: tonsil/ tonsilar foss (n=26), base of tongue (n=8)* cT-stage, n: TORS: T1 (n=17), T2 (n=17) RT: T1 (n=13), T2 (n=21) cN-stage, n: TORS: N0 (n=9), N1 (n=7), N2 (n=18) RT: N0 (n=12), N1 (n=5), N2 (n=17) Baseline ECOG status, n: TORS: score 0 (n=30), score 1 (n=4)	for patients deemed unfit for cisplatin could be modified, or cetuximab or weekly carboplatin could be administered. *** Out of 34 persons, 24 (71%) received radiotherapy, and 8 (24%) received chemotherapy (cisplatin: 5/8 (63%), carboplatin: 3/8 (38%), cetuximab: 0/8 (0%)). Median chemotherapy cycles: 6 (IQR:4.5-6) ***	Median chemotherapy cycles: 3 (IQR: 2-6) ***	Reasons (describe): No reasons described. However, from survival analysis: 1 person died and 3 were censored. From trial flow: 2 were lost to follow-up (unclear whether they were censored in the survival analyses or not)	RR (TORS/RT) = 8.00 (95%CI: 1.06-60.5) <u>All grades trismus, n (%):</u> TORS: 9/34 (23.5%) RT: 1/34 (2.9%) RR (TORS/RT) = 9.00 (95%CI: 1.21-67.21) **Taste:** <u>All taste alteration, n (%):</u> Only grade 1-2 alterations were observed. TORS: 17/34 (50%) RT: 19/34 (55.9%) RR (TORS/RT) = 0.89 (95%CI: 0.57-1.40) **Dryness of mouth:** <u>Grade 1-2 dry mouth, n (%):</u> TORS: 18/34 (52.9%) RT: 24/34 (70.6%) RR (TORS/RT) = 0.75 (95%CI: 0.51-1.10) <u>All grades dry mouth, n (%):</u> TORS: 18/34 (52.9%) RT: 25/34 (73.5%) RR (TORS/RT) = 0.72 (95%CI: 0.49-1.05) **Mucositis:** <u>Grade 1-2 oral mucositis, n (%):</u> TORS: 8/34 (23.5%)	

		RT: score 0 (n=30), score 1 (n=4) HPV p16-positive, n (%): TORS: 30 (88%) RT: 30 (88%) Groups comparable at baseline?				RT: 11/34 (32.4%) RR (TORS/RT) = 0.73 (95%CI: 0.33-1.58) <u>All grade oral mucositis, n (%):</u> TORS: 8/34 (23.5%) RT: 15/34 (44.1%) RR (TORS/RT) = 0.53 (95%CI: 0.26-1.09) **Quality of life:** MDADI: <u>MDADI total score, mean score (SD):</u> TORS: 80.1 (13) RT: 86.9 (11.4) Mean difference: 6.7 (95%CI: 0.2 to 13.2) Statistically significant difference between groups: p=0.042 <u>MDADI global sub-scale score, mean score (SD):</u> TORS: 79.3 (22.6) RT: 89.6 (15.1) Mean difference: 10.3 (95%CI: 0.2 to 20.4) Statistically significant difference between groups: p=0.046 <u>MDADI emotional sub-scale score, mean score (SD):</u> TORS: 81.3 (12.5) RT: 88.8 (12)	

						Mean difference: 7.4 (95%CI: 0.9 to 14) Statistically significant difference between groups: p=0.027 <u>MDADI functional sub-scale score, mean score (SD):</u> TORS: 86.5 (12) RT: 89.9 (11.5) Mean difference: 3.4 (95%CI: -2.9 to 9.6) No statistically significant difference between groups: p=0.28 <u>MDADI physical sub-scale score, mean score (SD):</u> TORS: 75.3 (16.5) RT: 83.1 (14.1) Mean difference: 7.9 (95%CI: -0.3 to 16) No statistically significant difference between groups: p=0.058 <u>MDADI composite score, mean score (SD):</u> TORS: 80.2 (13.1) RT: 86.7 (11.4) Mean difference: 6.5 (95%CI: 0.0 to 13.1) Statistically significant difference between groups: p=0.049 C30:	

						EORTC-C30 Global health status sub-scale score, mean score (SD): TORS: 77.9 (19.5) RT: 76.2 (20.9) No statistically significant difference between groups: p= 0.76 EORTC-C30 Physical functioning sub-scale score, mean score (SD): TORS: 9.4 (16.1) RT: 5.9 (7.2) No statistically significant difference between groups: p= 0.29 EORTC-C30 Role functioning sub-scale score, mean score (SD): TORS: 18.3 (30.1) RT: 11.1 (17.9) No statistically significant difference between groups: p= 0.27 EORTC-C30 Emotional functioning sub-scale score, mean score (SD): TORS: 14.9 (19.5) RT: 12.0 (15.9) No statistically significant difference between groups: p= 0.54	

						EORTC-C30 Cognitive functioning sub-scale score, mean score (SD): TORS: 13.8 (18.9) RT: 11.7 (15.9) No statistically significant difference between groups: p=0.66 EORTC-C30 Social functioning status sub-scale score, mean score (SD): TORS: 13.2 (20.1) RT: 6.4 (13.4) No statistically significant difference between groups: p= 0.14 EORTC-C30 Fatigue sub-scale score, mean score (SD): TORS: 18.1 (20.5) RT: 15.6 (13.5) No statistically significant difference between groups: p= 0.59 EORTC-C30 Nausea/ vomiting sub-scale score, mean score (SD): TORS: 5.0 (9.9) RT: 4.3 (10.9) No statistically significant difference between groups: p= 0.81 EORTC-C30 Pain sub-scale score, mean score (SD):	

						TORS: 21.8 (25.2) RT: 8.0 (16.3) Statistically significant difference between groups: p= 0.018 <u>EORTC-C30 Dyspnea sub-scale score, mean score (SD):</u> TORS: 7.8 (14.3) RT: 4.9 (12.1) No statistically significant difference between groups: p= 0.42 <u>EORTC-C30 Insomnia sub-scale score, mean score (SD):</u> TORS: 17.8 (27.3) RT: 28.4 (28.8) No statistically significant difference between groups: p= 0.16 <u>EORTC-C30 Appetite loss sub-scale score, mean score (SD):</u> TORS: 13.3 (27.1) RT: 16.0 (25.1) No statistically significant difference between groups: p= 0.70 <u>EORTC-C30 Constipation sub-scale score, mean score (SD):</u> TORS: 4.4 (14.5) RT: 8.6 (14.9) No statistically significant difference between groups: p= 0.29	

						EORTC-C30 Diarrhea sub-scale score, mean score (SD): TORS: 5.7 (15.6) RT: 2.5 (8.9) No statistically significant difference between groups: p= 0.34 EORTC-C30 Financial difficulties sub-scale score, mean score (SD): TORS: 14.9 (29.0) RT: 11.1 (24.5) No statistically significant difference between groups: p= 0.59 HN-35: EORTC-HN35 Pain sub-scale score, mean score (SD): TORS: 13.3 (14.9) RT: 9.0 (12.4) No statistically significant difference between groups: p= 0.23 EORTC-HN35 Swallowing sub-scale score, mean score (SD): TORS: 12.7 (16.1) RT: 7.4 (7.4) No statistically significant difference between groups: p= 0.11 EORTC-HN35 Senses sub-scale score, mean score (SD):	

						TORS: 20.6 (21.3) RT: 20.5 (22.8) No statistically significant difference between groups: p > 0.99 <u>EORTC-HN35 Speech sub-scale score, mean score (SD):</u> TORS: 7.7 (9.9) RT: 5.8 (9.9) No statistically significant difference between groups: p= 0.48 <u>EORTC-HN35 Social eating sub-scale score, mean score (SD):</u> TORS: 11.8 (14.4) RT: 7.1 (10.3) No statistically significant difference between groups: p= 0.16 <u>EORTC-HN35 Social contact sub-scale score, mean score (SD):</u> TORS: 4.6 (10.1) RT: 1.5 (6.5) No statistically significant difference between groups: p= 0.17 <u>EORTC-HN35 Less sexuality sub-scale score, mean score (SD):</u> TORS: 22.0 (27.6) RT: 17.3 (25.2) No statistically significant difference between groups: p= 0.51	

						EORTC-HN35 Teeth sub-scale score, mean score (SD): TORS: 12.2 (22.3) RT: 1.2 (6.4) Statistically significant difference between groups: p= 0.014 EORTC-HN35 Opening mouth sub-scale score, mean score (SD): TORS: 11.1 (22.0) RT: 6.4 (13.4) No statistically significant difference between groups: p= 0.33 EORTC-HN35 Dry mouth sub-scale score, mean score (SD): TORS: 44.4 (30.7) RT: 53.1 (31.0) No statistically significant difference between groups: p= 0.30 EORTC-HN35 Sticky saliva sub-scale score, mean score (SD): TORS: 31.1 (34.9) RT: 32.1 (28.5) No statistically significant difference between groups: p= 0.91 EORTC-HN35 Coughing sub-scale score, mean score (SD): TORS: 22.2 (23.7)	

						RT: 24.7 (25.5) No statistically significant difference between groups: p= 0.71 <u>EORTC-HN35 Felt ill sub-scale score, mean score (SD):</u> TORS: 6.7 (13.6) RT: 3.7 (10.7) No statistically significant difference between groups: p= 0.36 <u>EORTC-HN35 Pain killers sub-scale score, mean score (SD):</u> TORS: 44.8 (50.6) RT: 14.8 (36.2) Statistically significant difference between groups: p= 0.013 <u>EORTC-HN35 Nutritional supplements sub-scale score, mean score (SD):</u> TORS: 24.1 (43.5) RT: 29.6 (46.5) No statistically significant difference between groups: p= 0.65 <u>EORTC-HN35 feeding tube sub-scale score, mean score (SD):</u> TORS: 0.0 (0) RT: 3.7 (19.2) Statistically significant difference between groups: p= 0.33	

						<u>EORTC-HN35 weight loss sub-scale score, mean score (SD):</u> TORS: 20.7 (41.2) RT: 3.7 (19.2) No statistically significant difference between groups: p= 0.053 <u>EORTC-HN35 weight gain sub-scale score, mean score (SD):</u> TORS: 37.9 (49.4) RT: 40.7 (50.1) No statistically significant difference between groups: p= 0.83 NDII: <u>NDII total score, mean score (SD):</u> TORS: 81.5 (28.7) RT: 92.3 (10.0) No statistically significant difference between groups: p= 0.072 PNQ: <u>PNQ Numbness sub-scale score, mean score (SD):</u> TORS: 0.5 (0.8) RT: 0.3 (0.6) No statistically significant difference between groups: p= 0.57	

						PNQ Weakness sub-scale score, mean score (SD): TORS: 0.5 (1.0) RT: 0.2 (0.5) No statistically significant difference between groups: p= 0.14 VSI-10: VHI-10 total score, mean score (SD): TORS: 4.5 (4.3) RT: 4.4 (4.6) No statistically significant difference between groups: p= 0.89 FOIS: FOIS, level 7 reached at 1 year, n(%): TORS: 26/31 (84%) RT: 27/27 (100%) No statistically significant difference between groups: p= 0.055	

Notes:

1. **Prognostic balance between treatment groups is usually guaranteed in randomized studies, but non-randomized (observational) studies require matching of patients between treatment groups (case-control studies) or multivariate adjustment for prognostic factors (confounders) (cohort studies); the evidence table should contain sufficient details on these procedures.**
2. **Provide data per treatment group on the most important prognostic factors ((potential) confounders).**
3. **For case-control studies, provide sufficient detail on the procedure used to match cases and controls.**
4. **For cohort studies, provide sufficient detail on the (multivariate) analyses used to adjust for (potential) confounders.**

Table of excluded studies

Author and year	Reason for exclusion
Beitler 2016	Delphi consensus study
Channir 2018	Article in Danish
De Almeida 2014	No RCT included
Dhanireddy 2019	Non-randomized design
Dias 2017	Narrative review
Fabian 2019	Article in German
Genden 2011	Non-randomized design
Howard 2014	No RCT included
Ling 2016	Non-randomized design
Monnier 2015	Opinion statement
More 2013	Non-randomized design
O'Hara 2016	Non-randomized design
Rubek 2017	Non-randomized design
Rudmik 2015	Cost-effectiveness study
Sharma 2016	Non-randomized design
Smith 2015	Non-randomized design
Thuy 2018	Non-randomized design
Viros Porcuna 2019	Non-randomized design
Wang 2015	No RCT included
Yeah 2015	No RCT included

Literature search strategy

Uitgangsvraag: HHT, behandeling orofarynxcarcinoom (T1-T2 N0 N1)	
Database(s): Ovid/Medline, Embase	Datum: 6-1-2019
Periode:	Talen:
Toelichting: Het aantal referenties dat wordt gevonden is niet groot vandaar dat er geen beperkingen op tijd zijn doorgevoerd.	

	Inclusief dubbele referenties	Ontdubbeld
SR	41	32
RCT	171	132
Observationeel		
Totaal		

Zoekverantwoording

PubMed

History and Search Details

Download

Delete

Search	Query	Results
#15	Search: #10 AND #12	65
#13	Search: #10 AND #11	13
#12	Search: randomized controlled trial(pt) OR controlled clinical trial(pt) OR randomized(tiab) OR placebo(tiab) OR drug therapy(sh) OR randomly(tiab) OR trial(tiab) OR groups(tiab)	4,663,898
#11	Search: Search ("Meta-Analysis as Topic"(Mesh) OR “Meta-Analysis”(Publication Type) OR metaanaly*(tiab) OR metanaly*(tiab) OR meta-analy*(tiab) OR meta synthes*(tiab) OR metasynthes*(tiab) OR meta ethnograph*(tiab) OR metaethnograph*(tiab) OR meta summar*(tiab) OR metasummar*(tiab) OR meta-aggregation(tiab) OR metareview(tiab) OR meta-review(tiab) OR overview of	161,726

	reviews(tiab) OR ((systematic*(ti) OR scoping(ti) OR umbrella(ti) OR meta-narrative(ti) OR metanarrative(ti) OR evidence based(ti)) AND (review*(ti) OR overview*(ti))) OR ((evidence(ti) OR narrative(ti) OR metanarrative(ti) OR qualitative(ti)) AND synthesis(ti)) OR systematic review(pt) OR prisma(tiab) OR preferred reporting items(tiab) OR quadas*(tiab) OR systematic review*(tiab) OR systematic literature(tiab) OR structured literature search(tiab) OR systematic overview*(tiab) OR scoping review*(tiab) OR umbrella review*(tiab) OR mapping review*(tiab) OR systematic mapping(tiab) OR evidence synthes*(tiab) OR narrative synthesis(tiab) OR metanarrative synthesis(tiab) OR research synthesis(tiab) OR qualitative synthesis(tiab) OR realist synthesis(tiab) OR realist review(tiab) OR realist evaluation(tiab) OR systematic qualitative review(tiab) OR mixed studies review(tiab) OR mixed methods synthesis(tiab) OR mixed research synthesis(tiab) OR quantitative literature review(tiab) OR systematic evidence review(tiab) OR evidence-based review(tiab) OR comprehensive literature search(tiab) OR integrated review*(tiab) OR integrated literature review(tiab) OR integrative review*(tiab) OR integrative literature review*(tiab) OR structured literature review*(tiab) OR systematic search and review(tiab) OR meta-narrative review*(tiab) OR metanarrative review(tiab) OR systematic narrative review(tiab) OR systemic review(tiab) OR systematized review(tiab) OR systematic research synthesis(tiab) OR bibliographic*(tiab) OR hand-search*(tiab) OR handsearch*(tiab) OR manual search*(tiab) OR searched manually(tiab) OR manually searched(tiab) OR journal database*(tiab) OR review authors independently(tiab) OR reviewers independently(tiab) OR independent reviewers(tiab) OR independent review authors(tiab) OR electronic database search*(tiab) OR (study selection(tiab) AND data extraction(tiab)) OR (selection criteria(tiab) AND data collection(tiab)) OR (selection criteria(tiab) AND data analysis(tiab)) OR (evidence acquisition(tiab) AND evidence synthesis(tiab))) NOT ("Comment" (Publication Type) OR "Letter" (Publication Type)) NOT (“Animals”(Mesh) NOT “Humans”(Mesh))	
#10	Search: #6 AND #9	366
#9	Search: "Robotic Surgical Procedures"(MeSH Terms) OR "Microsurgery"(MeSH Terms) OR robot assisted surg*(tiab) OR tors(tiab) OR microsurg*(tiab)	54,039
#6	Search: "Oropharyngeal Neoplasms"(Mesh) OR "oropharyngeal tumor"(tiab) OR "oropharyngeal tumour"(tiab) OR "oropharyngeal tumours"(tiab) OR "oropharynx tumors"(tiab) OR "oropharyngeal cancer"(tiab) OR "oropharyngeal cancers"(tiab) OR "oropharynx cancer"(tiab) OR "oropharynx cancers"(tiab) OR "oropharyngeal carcinoma"(tiab) OR "oropharyngeal carcinomas"(tiab) OR "oropharynx carcinoma"(tiab) OR "oropharynx carcinomas"(tiab) OR "oropharyngeal neoplasm"(tiab) OR "oropharyngeal neoplasms"(tiab) OR "oropharynx neoplasm"(tiab) OR "oropharynx neoplasms"(tiab)	10,135

Embase Session Results (6 Jan 2020)

No.	Query	Results
#15	#12 AND #13	**106**
#14	#11 AND #13	**28**
#13	#3 NOT ('conference abstract'/it OR 'conference paper'/it OR 'editorial'/it OR 'letter'/it OR 'note'/it) NOT (('animal experiment'/exp OR 'animal model'/exp OR 'nonhuman'/exp) NOT 'human'/exp)	**455**
#12	('clinical trial'/exp OR 'randomization'/exp OR 'single blind procedure'/exp OR 'double blind procedure'/exp OR 'crossover procedure'/exp OR 'placebo'/exp OR 'prospective study'/exp	**2331808**

	OR rct:ab,ti OR random*:ab,ti OR 'single blind':ab,ti OR 'randomised controlled trial':ab,ti OR 'randomized controlled trial'/exp OR placebo*:ab,ti) NOT 'conference abstract':it	
#11	('meta analysis'/de OR cochrane:ab OR embase:ab OR psycinfo:ab OR cinahl:ab OR medline:ab OR ((systematic NEAR/1 (review OR overview)):ab,ti) OR ((meta NEAR/1 analy*):ab,ti) OR metaanalys*:ab,ti OR 'data extraction':ab OR cochrane:jt OR 'systematic review'/de) NOT (('animal experiment'/exp OR 'animal model'/exp OR 'nonhuman'/exp) NOT 'human'/exp)	**470833**
#10	#8 NOT #3	**1**
#9	#3 AND #8	**3**
#8	#4 OR #5 OR #6 OR #7	**4**
#7	27531879	**1**
#6	31416685	**1**
#5	26461255	**1**
#4	masterson AND minimally AND invasive AND surgery AND versus AND radiotherapy AND 2016	**1**
#3	#1 AND #2	**662**
#2	'microsurgery'/exp OR 'robot assisted surgery'/exp OR 'robot assisted surg*':ti,ab,kw OR tors:ti,ab,kw OR microsurg*:ti,ab,kw OR 'transoral robotic surgery'/exp	**59879**
#1	'oropharynx cancer'/exp OR ((oropharyn* NEAR/3 (tumor* OR tumour* OR cancer* OR carcinoma* OR neoplasm*)):ti,ab,kw)	**14616**

Module 12.4 Behandeling HPV-positieve orofarynx tumoren

Uitgangsvraag

Dienen HPV-positieve orofarynx tumoren op een andere wijze behandeld te worden dan HPV-negatieve orofarynx tumoren? Is er plaats voor de-escalatie?

Introduction

Standard treatment with cisplatin-based chemoradiotherapy for stage III-IV human papillomavirus (HPV)-positive oropharyngeal cancer results in considerable acute and long-term toxicity. Wide consensus exists about the need for de-escalation treatments with decreased toxicity and similar survival. Forms of de-escalation in this setting are cetuximab-based bioradiotherapy, radiotherapy alone, lowering the dose of the radiotherapy, low dose cisplatin and minimally invasive transoral surgery followed by de-intensified adjuvant therapy compared to standard dose cisplatin based concurrent chemoradiotherapy.

Search and select

A systematic review of the literature was performed to answer the following question:

> What are the effects of a de-escalating strategy in the treatment of advanced human papilloma virus (HPV)-associated oropharyngeal cancer (OPC) compared to care-as-usual for OPC?

P: patients with HPV-positive oropharyngeal cancer;
I: de-escalating strategy, such as cetuximab-based bioradiotherapy, radiotherapy alone or lowering the dose of the radiotherapy;
C: **c**are-as-usual, such as cisplatin-based chemoradiotherapy;
O: overall survival, complications/adverse events, quality of life, head and neck cancer–specific functional outcomes.

<u>Relevant outcome measures</u>

The guideline development group considered overall survival and quality of life as critical outcome measures for decision making; and complications/adverse events, functional outcomes and (work) participation as important outcome measures for decision making.

The working group defined the outcome measures as follows:

Overall survival:	Overall survival (defined as time from randomisation to death from any cause) after a minimum follow-up of 3 years.
Progression free survival:	Progression free survival (time during and after the treatment of a disease that the patient lived with the disease but it did not get worse) after a minimum follow-up of 3 years.
Quality of life:	Quality of life (overall or regarding a specific domain) as measured with a validated and reliable instrument such as the SF-36 or European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ).
Complications/adverse events:	All negative effects related to the treatment (lethal, acute/serious, chronic).
Functional outcomes:	Swallowing, oral pain, dry mouth, dental health, opening mouth/trismus, taste, excess/thick mucous/saliva, shoulder disability/ motion assessed by a validated and reliable instrument (Chera, 2014).

(Work) Participation: Participation in school, work and/or informal care.

The working group defined a minimal clinically relevant difference as *(in line with "NVMO-commissie ter Beoordeling van Oncologische Middelen (BOM)")*:

- > 5% difference or more > 3% and HR<0.7 in overall survival.
- HR< 0.7 for progression free survival.

And, in case of absence of a clinically relevant difference in overall survival of progression free survival:

- A minimal clinically important difference of 10 points on the quality of life instrument EORTC QLQ-C30 (in line with Mehanna, 2019) or a difference of a similar magnitude on other quality of life instruments.
- Statistically significant less complications/adverse events.
- Statistically significant better functional outcomes, work participation.

Search and select (Methods)

The databases Medline (via OVID) and Embase were searched with relevant search terms until November 2019. The detailed search strategy is depicted under the tab Methods. The systematic literature search resulted in 409 hits (45 found with a filter for SRs, 364 found with a filter for RCTs). Only randomized controlled trials (whether or not included in systematic reviews) were considered for inclusion.

Systematic reviews

First, SRs were selected based on the following criteria: the population consists (mostly) of patients with HPV-positive oropharyngeal cancer, the study compares one or more de-escalating treatment modalities with care-as-usual, the study is a systematic review (SR) and describes one or more of the selected outcome measures.

Nineteen SRs were initially selected based on title and abstract screening. After reading the full text, 15 SRs were excluded (see the table with reasons for exclusion under the tab Evidence tables), and 4 SRs were included: Materson (2014) (comparing all de-escalation treatments); Suton 2019 (comparing cetuximab-based bioradiotherapy versus cisplatin); Szturz (2017) (comparing low dose versus high dose cisplatin) and Howard (2018) (comparing minimally invasive transoral surgery followed by de-intensified adjuvant therapy (either omission of chemotherapy or reduced-dose radiotherapy) versus minimally invasive transoral surgery followed by standard concurrent chemoradiotherapy or standard-dose radiotherapy).

Randomized controlled trials

Based on the SRs found, it was decided to select randomized controlled trials (RCTs) based on the following criteria: published after 2016, the population consists (mostly) of patients with HPV-positive oropharyngeal cancer, the study compares one or more de-escalating treatment modalities with care-as-usual, the study is a RCT and describes one or more of the selected outcome measures.

Ten articles were initially selected based on title and abstract screening. After reading the full text, 9 were excluded (see the table with reasons for exclusion under the tab Evidence tables), and 1 RCT was included: Misiukiewicz 2019 (comparing reduced dose chemoradiation with standard-dose chemoradiotherapy).

Results

Literature per type of de-escalating strategies

In 2014, Materson searched for all de-escalation treatment studies for HPV-associated, locally advanced (stage III-IV) oropharyngeal squamous cell carcinoma in their Cochrane literature review. They found no RCTs. In 2019, this review has not been updated yet. The most recent review on all de-escalation treatment strategies found in the present search for this guideline module was written by Stock in 2018. This was a narrative review, and not a systematic review, and was therefore not included.

We found the following literature on specific de-escalation strategies (Table 12.4):

1. Cetuximab-based bioradiotherapy versus cisplatin-based chemoradiotherapy (Suton, 2019 search up to December 2018).
2. Lowering the dose of the chemotherapy:
   a. Low dose versus high dose cisplatin (Szturz (2017) found no RCTs; search up to 2015).
3. Radiotherapy alone or lowering the dose of the radiotherapy:
   a. Radiotherapy alone (no recent SRs, Stock 2018 reported only retrospective analyses).
   b. Low dose cisplatin with concurrent lower dose radiotherapy (5600 cGy with weekly carboplatin) versus low dose cisplatin with concurrent standard-dose radiotherapy (7000 cGy with weekly carboplatin) after induction chemotherapy response (Misiukiewicz, 2019).
4. Minimally invasive transoral surgery followed by de-intensified adjuvant therapy (either omission of chemotherapy or reduced-dose radiotherapy) compared to minimally invasive transoral surgery followed by standard concurrent chemoradiotherapy or standard-dose radiotherapy (Howard 2018 found in their Cochrane review no RCTs, search up to April 2018).

Table 12.4 Literature on specific de-escalation strategies

De-escalation strategy	SR	Included RCTs
Cetuximab-based bioradiotherapy (versus cisplatin-based chemoradiotherapy)	Suton, 2019, search up to December 2018)	RTOG 1016 (Gillison, 2019) De-ESCALaTE (Mehanna, 2019)
Low dose cisplatin (versus high dose)	Szturz, 2017, search up to 2015	none
Radiotherapy alone	no recent SR	none
Reduced radiation dose chemoradiation (5600 cGy with weekly carboplatin) (versusstandard-dose chemoradiotherapy (7000 cGy with weekly carboplatin))		Quarterback trial (Misiukiewicz, 2019)
Minimally invasive transoral surgery followed by de-intensified adjuvant therapy	Howard, 2018, search up to April 2018	none

Reduced dose chemoradiation

The study of Misiukiewicz (2019) was terminated early due to lack of financial support for infrastructure to expand to multiple sites. Only 20 patients in total were included: 12 in the standard chemotherapy dose and reduced radiation dose group and 8 in the standard-dose chemoradiotherapy group. We decided that this very small sample size did not allow us to draw any useful conclusions for the purpose of developing a guideline. Details of the study are presented in Table 12.52a, but the study was excluded from the summary of the literature.

Table 12.5a. Details of the trial that compared reduced dose chemoradiation with standard-dose chemoradiotherapy

Trial	Intervention (n, treatment)	Control (n, treatment)	Outcomes
Quarterback trial (Misiukiewicz, 2019)	n=12 3 cycles of IC with docetaxel, cisplatin and fluorouracil (TPF) Clinical responders who were HPV positive by type-specific PCR received standard chemo and reduced dose radiotherapy (5600 cGy) with weekly carboplatin.	n=8 3 cycles of IC with docetaxel, cisplatin and fluorouracil (TPF) Clinical responders who were HPV positive by type-specific PCR received standard-dose chemoradiotherapy (7000 cGy) with weekly carboplatin.	• 3-yr overall survival • 3-yr progression-free survival • Toxicity

Cetuximab-based bioradiotherapy

Two trials were included in the summary of the literature (Gillison, 2019; Mehanna, 2019). Important study characteristics and results are summarized in the evidence tables. The assessment of the risk of bias is summarized in the risk of bias tables.

Summary of literature

Cetuximab-based bioradiotherapy versus cisplatin

Two RCTs (RTOG 1016 and De-ESCALaTE) directly compared the efficacy of cisplatin (CDDP) versus cetuximab (C225) given concurrently with RT as definitive treatment of p16-positive, non-metastatic, and locally advanced/unresectable OPC. In both studies p16-positive immunohistochemical staining was used as surrogate marker for HPV-positivity and both studies included patients with advanced stage tumors. Gillison 2019 published the results after a median follow-up duration of 4.5 years of the RTOG 1016 non-inferiority trial. The sample consisted of T1 (n=175), T2 (n=325), T3 (n=208), and T4 (n=97) tumors, and N0 (n=34), N1 (n=45), N2a (n=115), N2b (n=417), N2c (n=165), and N3 (n=29) lymph node metastases. Mehanna (2019) published the results of the De-ESCALaTE trial at 24 months. The sample consisted of T1-2 (n=216) and T3-4 (n=118) tumors, and N0-1 (n=81) and N2-3 (n=253) lymph node metastases. Details about these two trials are presented in Table 12.5b.

Table 12.5b. Details of the trials that compared cetuximab-based bioradiotherapy versus cisplatin

Trial	Intervention (n, treatment)	Control (n, treatment)	Outcomes
RTOG 1016 (Gillison, 2019)	n=399 Intravenous cetuximab at a loading dose of 400 mg/m² 5–7 days before radiotherapy initiation, followed by cetuximab 250 mg/m² weekly for seven doses (total 2150 mg/m²). All patients received accelerated intensity-modulated radiotherapy delivered at 70 Gy in 35 fractions over 6 weeks at six fractions per week (with two fractions given on one day, at least 6 h apart).	n=406 Cisplatin (commercially available and obtained by each individual institution) 100 mg/m² on days 1 and 22 of radiotherapy (total 200 mg/m²) All patients received accelerated intensity-modulated radiotherapy delivered at 70 Gy in 35 fractions over 6 weeks at six fractions per week (with two fractions given on one day, at least 6 h apart).	• 5-yr overall survival • 5-yr progression-free survival • Toxicity • Swallowing • Dental health
De-ESCALaTE (Mehanna, 2019)	N=168 Intravenous **cetuximab** 400 mg/m² loading dose 1 week before	N=166 Three doses of intravenous **cisplatin** 100 mg/m² on days 1, 22, and 43 of radiotherapy.	• 2-yr overall survival • Toxicity • Quality of life • Swallowing

	followed by seven weekly infusions of 250 mg/m² during radiotherapy All patients received accelerated intensity-modulated radiotherapy delivered at 70 Gy in 35 fractions over 7 weeks at five fractions per week.	All patients received accelerated intensity-modulated radiotherapy delivered at 70 Gy in 35 fractions over 7 weeks at five fractions per week. NB: only 38% of the patients in the cisplatinum group were able to get all three courses.	

Important remarks and differences between the studies

The studies differed in some aspects: low risk HPV-positive oropharyngeal cancer (Mehanna, 2019) versus low and moderate risk HPV-positive oropharyngeal cancer (Gillison, 2019), conventional (Mehanna 2019) versus accelerated radiotherapy (Gillison, 2019), three (Mehanna, 2019) versus two courses of cisplatin (Gillison, 2019) and toxicity (Mehanna 2019) versus survival (Gillison, 2019) as a primary outcome measure.

In the De-ESCALaTe study, 6% of the p16 positive tumors in HPV DNA in situ hybridization were negative, so that the difference between the two groups could be even greater. For a more in-depth discussion of the two studies, we refer to De Bree and Devriese 2019.

Results

Overall survival (follow-up ≥ 3 years)

In November 2019 only long-term follow-up data of the RTOG 1016 trial (Gillison, 2019) were available. In this trial 133 patients died after a median follow-up duration of 4.5 years: 78 (78/399, 19.5%) in the cetuximab group and 55 (55/406, 13.5%) in the cisplatin group (HR 1.45, one-sided 95% upper CI 1.94; p=0.5056 for non-inferiority; RR 1.44, 95% CI 1.05 to 1.98). The boundary for clinical relevance (>5% difference) was exceeded, indicating worse overall survival after treatment with cetuximab.

Progression free survival

Five-year progression free-survival was 67.3% (95% CI 62.4 to 72.2) in the cetuximab group and 78.4% (73.8 to 83.0) in the cisplatin group (HR 1.72, 95% CI 1.29 to 2.29; p=0.0002) (Gillison, 2019). The boundary for clinical relevance was exceeded, indicating better progression free survival after treatment with cisplatin.

Mehanna (2019) did not report progression free survival as such, but reported that they observed:

- “Significantly fewer recurrences with cisplatin than with cetuximab (ten (6%) versus 29 (18%); log-rank p=0.0007).
- Significantly fewer locoregional recurrences with cisplatin than with cetuximab (3% versus 12%, log-rank p=0.0026).
- Significantly fewer distant metastases with cisplatin than with cetuximab (3% versus 9%, log-rank p=0.0092).
- Five (3%) patients in each group developed second primaries.”

Toxicity

Mehanna (2019) reported the mean number of acute, late, and overall toxicity events per patient. No statically significant differences were found. Details are presented in Table 12.6.

Table 12.6 Mean number of acute, late, and overall toxicity events per patient, by treatment group (Mehanna, 2019)

	mean number of	mean number of

	events per patient in cisplatin group	events per patient in cetuximab group
Severe short-term toxicities	4.4 (95% CI 3.9–4.97)	4.4 (3.8–4.9)
All-grade short-term toxicity	20.0 (95% CI 18.8–21.1	20.4 (19.2–21.5)
Severe late toxicity events	0.4 (95% CI 0.29-0.54)	0.5 (0.030-0.67)
All-grade late toxicity events	9.4 (95% CI 8.53–10.34)	9.9 (95% CI 9.02-10.72)

Similar to Mehanna (2019), Gillison (2019) found that with regard to late toxicity in the cetuximab versus cisplatin groups, no statistically significant differences were found. Gillison (2019) found, however, that patients in the cetuximab group had a significantly lower mean number of grade 3 to 4 acute toxicity events per patient than did those in the cisplatin group (raw T-score 2.35 versus 3.19; p<0·0001), corresponding to a 40% lower acute toxicity burden.

Adverse events
Mehanna (2019) found that there were significantly more serious adverse events with cisplatin than with cetuximab. 162 adverse events (mean rate of one event per patient) occurred in patients receiving cisplatin and 95 events (mean rate of 0.6 events per patient) occurred in patients receiving cetuximab (p<0.0001).

Gillison (2019) reported that the proportion of one or more grade 3 to 4 acute adverse events was similar in the cetuximab and cisplatin groups (305 of 394 patients, 77.4%, 95% CI 73.0 to 81.5 versus 325 of 398 patients, 81.7%, 77.5 to 85.3; p=0.16).

Next to this, Gillison (2019) reported that the number of early deaths (death due to adverse event or within 30 days of treatment completion) was the same in the cetuximab and cisplatin groups (6 of 394 patients in the cetuximab group; 6 of 398 in the cisplatin group; 1.5%, 95% CI 0.6 to 3.3; p=1.0).

Quality of life
Mehanna (2019) did not find statistically significant differences in the mean global quality-of-life score on EORTC QLQ-C30 between treatment groups at any of the timepoints. A statistically significant difference in social functioning was observed in favour of cetuximab at the end of treatment (mean difference of 8.67 points, p=0.0374), but this difference disappeared 6 months later. At 12 months and 24 months, a significant difference in role functioning was observed in favour of cisplatin (difference in mean scores of 8.32 points, p=0.0173. None of the differences reached the minimal clinically important difference of 10 points.

Gillison (2019) reported only the EORTC QLQ-H&N35 swallowing domain (see below: Functional outcomes). Additional quality of life endpoints will be reported in future publications.

Functional outcomes, work participation.
Gillison (2019) found that patient-reported severity of swallowing problems, as measured with the EORTC QLQ-H&N35 subscale, increased in both the cetuximab and cisplatin groups from pretreatment to end of treatment, but no difference was observed between groups in change scores from baseline (mean 47.4 versus 48.0; p=0.86). At 1 year, the cetuximab group had a statistically significant increase in symptoms from pretreatment compared with the cisplatin group (7.6 versus 2.5; p=0.04), but this difference was below the, by the authors of the study, estimated clinically important difference.

None of the studies reported on (work) participation.

Level of evidence of the literature
The levels of evidence regarding the outcome measures overall survival, progression free survival, quality of life and functional outcomes were downgraded by one level because of study limitations (risk of bias in only one study) and two levels because of the number of events in only one study (very serious imprecision, two levels). The level of evidence regarding the outcome measures complications/adverse events and toxicity were downgraded by one level because of study limitations (risk of bias), one level because of the low number of events (serious imprecision) and one level because of inconsistency (conflicting results).

Conclusions
What are the effects of a de-escalating strategy in the treatment of human papilloma virus (HPV)-associated oropharyngeal cancer (OPC) compared to care-as-usual?

Radiotherapy alone or lowering the dose of the radiotherapy

- **GRADE**	We found no credible literature on de-escalating strategies, such as radiotherapy alone or lowering the dose of the radiotherapy.

Cetuximab-based bioradiotherapy versus cisplatin-based chemoradiotherapy

Very low **GRADE**	**Overall survival** may be lower after treatment with the de-escalating strategy "cetuximab-based bioradiotherapy", but the evidence is very uncertain. *Sources: (Gillison, 2019)*

Very low **GRADE**	**Progression free survival** may be lower after treatment with the de-escalating strategy "cetuximab-based bioradiotherapy", but the evidence is very uncertain. *Sources: (Gillison, 2019)*

Very low **GRADE**	There may be no difference in **complications/adverse events and toxicity,** but the evidence is very uncertain. *Sources: (Gillison, 2019, Mehanna, 2019)*

Very low **GRADE**	The evidence is very uncertain about the effect of the de-escalating strategy "cetuximab-based bioradiotherapy" on **quality of life**. *Sources: (Mehanna, 2019)*

Very low **GRADE**	The evidence is very uncertain about the effect of the de-escalating strategy "cetuximab-based bioradiotherapy" **on swallowing problems**. Other functional outcomes or work participation were not reported in the studies. *Sources: (Gillison, 2019; Mehanna, 2019)*

Overwegingen - van bewijs naar aanbeveling
De werkgroep is van mening dat, op basis van het anno 2019 beschikbare bewijs over de-escalatie strategieën, er geen reden is om HPV-positieve tumoren op een andere wijze te behandelen dan HPV-negatieve tumoren. Studies met zeer lage bewijskracht laten een

lagere overleving zien in de de-escalatie groep. Volgens de werkgroep is er alleen binnen studieverband plaats voor de-escalatie van de behandeling.

Voor- en nadelen van de interventie en de kwaliteit van het bewijs
Alhoewel de-escalatie van de behandeling van HPV positieve orofarynxtumoren in meerdere studies is onderzocht, zijn er op dit moment slechts twee volledig uitgevoerde, gerandomiseerde studies beschikbaar. In deze studies werd concurrente radiotherapie met cetuximab vergeleken met concurrente radiotherapie met cisplatin. Van slechts één studie is op dit moment de lange-termijn follow-up data (> 3 jaar) beschikbaar (Gillison, 2019). De bewijskracht van deze ene studie is zeer laag. We zijn daarom onzeker of behandeling met cetuximab daadwerkelijk leidt tot een lagere overleving (6%-punt lager) en lagere "progression-free survival" (11%-punt lager).

Alhoewel de 2-jaars follow-up data van Mehanna (2019) niet voldoet aan de definitie van "lange-termijn follow-up", laat ook deze data een lagere overleving zien in de cetuximabgroep (20/168, 11.9%) vergeleken met de cisplatingroep (6/166, 3.6%).

De anno 2019 beschikbare data laat geen verschil zien in complicaties, toxiciteit, kwaliteit van leven of functionele uitkomsten, zoals slikproblemen. De bewijskracht is zeer laag. We kunnen daarom niet met zekerheid stellen dat er daadwerkelijk geen verschil is.

De werkgroep is van mening dat, op basis van het anno 2019 beschikbare bewijs over chemoradiotherapie met cetuximab vergeleken met chemoradiotherapie met cisplatin, er geen reden is om HPV-positieve tumoren op een andere wijze te behandelen dan HPV-negatieve tumoren.

Na de zoekdatum in deze richtlijnmodule werden er nog twee RCTs geïdentificeerd (Gebre-Medhin, 2020; Yom, 2021).

Gebre-Medhin (2020) rekruteerde patiënten met stadium III-IV (volgens de UICC TNM7-classificering) orofarynx-, hypofarynx-, mondholte of larynxcarcinomen zonder metastasen op afstand die in aanmerking kwamen voor een curatieve behandeling met radiotherapie. Patiënten werden gerandomiseerd en ontvingen 400 mg/m^2 intraveneuze cetuximab een week voor aanvang van de radiotherapie en vervolgens 250 mg/m^2 per week (n=149) óf ontvingen wekelijks 40 mg/m^2 intaveneuze cisplatin tijdens radiotherapie (n=149). In beide behandelgroepen werden patiënten geëxcludeerd (door middel van screening vóór het verstrekken van de behandeling) of vielen patiënten uit (vanwege sterfte of het stoppen van de behandeling) waardoor er bij n=144 patiënten een tumor respons evaluatie kon plaats vinden in de cetuximab-groep, tegenover n=143 in de cisplatin-groep. De tumorlocatie in de cetuximab-groep (n=125 orofarynx, n=8 mondholte, n=6 larynx, n=7 hypofarynx) en in de cisplatin-groep (n=123 orofarynx, n=7 mondholte, n=6 larynx, n=9 hypofarynx) leken gelijk tussen de groepen te zijn verdeeld. De radiotherapie had conventionele fractionering, maar patiënten met een T3-4-stadium carcinoom werden verder gerandomiseerd en ontvingen een radiotherapie dosis van 68Gy óf 73,1Gy. De patiënten werden aan de hand van p16 getest op de HPV-status. In de cetuximab-groep waren 108 patiënten (72,5%, n=14 negatief, n=1 missing) HPV-positief volgens p16 analyse, tegenover 113 patiënten (75,9%, n=11 negatief, n=1 missing) in de cisplatin groep. De auteurs vonden een niet-significant verschil in de algehele overleving op drie jaar follow-up in het voordeel van cisplatin. Daarnaast vonden de auteurs een statistisch significant verschil op de cumulatieve incidentie van het locoregionale falen op drie jaar in het voordeel van cisplatin. De auteurs rapporteerden ook dat de cumulatieve incidentie van het falen op afstand niet verschilde tussen de groepen. De

dosisescalatie van radiotherapie bij T3-4-stadium carcinomen zorgde niet voor een verbeterde lokale controle. De auteurs concludeerden dat cetuximab als concomitante behandeling met radiotherapie inferieur is aan cisplatin voor locoregionale controle en dat wellicht nieuwe studies nodig zijn om subgroepen te kunnen identificeren die mogelijk baat zouden hebben bij cetuximab als concomitante behandeling (Gebre-Medhin, 2020).

Yom (2021) rekruteerde patiënten met histologisch aangetoonde orofaryngeale plaveiselcelcarcinomen en een Zubrod-status van 0 of 1. Deelnemers hadden HPV-positieve T1-2N1-2bM0 of T3N0-2bM0 tumoren volgens de 7^{e} editie van het TNM-stadiëringssysteem en een rookgeschiedenis van ≤10 pakjesjaren. Hematologische en nier- en leverfuncties moesten adequaat zijn voor het toedienen van cisplatin. Patiënten werden gerandomiseerd en 152 deelnemers ontvingen intensiteitsgemoduleerde radiotherapie met concurrente cisplatin (60Gy in 30 fracties, 5 fracties per week, wekelijks 40 mg/m^2 cisplatin) tegenover 147 deelnemers die alléén intensiteitsgemoduleerde radiotherapie ontvingen (60Gy, 6 fracties per week). Voor een intermediair risico volume rond de primaire locatie, de betrokken halsniveaus en de direct naastgelegen niet-betrokken halsniveaus werd 54Gy voorgeschreven. Voor de overige electief behandelde halsniveaus werd 48Gy voorgeschreven. In de cisplatin groep zaten 133 mannen en 25 vrouwen met tumoren op de volgende locaties: orofarynx (n=5, waarvan n=1 faryngeale orofarynx), tonsil of fossa tonsilaris (n=83), basis van de tong (n=68), en de posterieure faryngeale wand (n=1). Het T-stadium in deze groep was T1 (n-64), T2 (n=67) of T3 (n=26) en het N-stadium was N0 (n=6), N1 (n=28), N2a (n=24) of N2b (n=99). In de groep die alléén radiotherapie ontving zaten 124 mannen en 25 vrouwen met tumoren op de volgende locaties: orofarynx (n=13), tonsil of fossa tonsilaris (n=78) of de basis van de tong (n=58). Het T-stadium was T1 (n=51), T2 (n=80) of T3 (n=18). Het N-stadium was N0 (n=7), N1 (n=34), N2a (n=19) of N2b (n=89). Er werden geen significante verschillen gevonden tussen de groepen voor algehele overleving, progressievrije overleving, het ontstaan van afstandsmetastasen en voor de verschilscore tussen baseline en een jaar later van de MDADI composietscore. Er werd een significant verschil gevonden voor het lokaal-regionaal falen, in het voordeel van de cisplatin-groep. Ook werden er meer graad 3-4 complicaties geobserveerd in de cisplatin-groep. De auteurs vermeldden dat de hoeveelheid late complicaties niet significant verschilden tussen de groepen. Er werd door de auteurs geconcludeerd dat de resultaten voldoende rechtvaardiging geven voor verder onderzoek in een fase III trial.

Andere de-escalatie strategieën, zoals behandeling met alleen radiotherapie of het verlagen van de radiotherapie of chemotherapie dosis, zijn niet in gerandomiseerd onderzoek onderzocht.

<u>Waarden en voorkeuren van patiënten (en eventueel hun verzorgers)</u>
Windon (2019) onderzocht in een prospectieve studie de behandelingsdoelen van patiënten met HPV-positieve orofarynxtumoren voorafgaand aan hun behandeling en na hun behandeling. Patiënten (n=37) vulden tweemaal (bij diagnosestelling en gemiddeld 10 maanden later) een vragenlijst in over hun behandelingsdoelen en zorgen. Uit deze studie bleken behandelingsdoelen voor het grootste deel onveranderd. Patiënten vonden genezing van de kanker, overleving en slikken zowel voor als na de behandeling het belangrijkst. De enige significante verandering was dat na de behandeling patiënten “een vochtige mond hebben” belangrijker vonden dan voorafgaand aan de behandeling.

Op basis van de bevindingen van Windon (2019) stelt de werkgroep dat zij het gerechtvaardigd vinden om bij het formuleren van de aanbeveling ervan uit te gaan dat in

het algemeen patiënten de behandelingsoptie met de beste kans op genezing en overleving zullen prefereren.

Kosten (middelenbeslag)
De werkgroep heeft geen informatie gevonden over de kosteneffectiviteit van de-escalatie strategieën. De werkgroep heeft dit aspect daarom niet meegewogen bij het formuleren van de aanbeveling. De werkgroep verwacht geen relevante impact op de zorgkosten door de aanbeveling.

Aanvaardbaarheid, haalbaarheid en implementatie
De werkgroep is van mening dat de aanbeveling aanvaardbaar is voor zowel zorgverleners als patiënten. De werkgroep verwacht dat het uitvoeren van de aanbeveling haalbaar is en implementeerbaar. De aanbeveling sluit aan bij de huidige werkwijze in de praktijk.

Aanbeveling

De-escaleer de behandeling van patiënten met een HPV-positief orofarynxcarcinoom alleen in studieverband.

Literatuur

De Bree R and Devriese L.A. Radiotherapie met cisplatinum of cetuximab voor de behandeling van humaan papillomavirus-positieve orofarynxcarcinomen. Ned Tijdschr Oncol 2019;16:79-81.

Chera BS, Eisbruch A, Murphy BA, Ridge JA, Gavin P, Reeve BB, Bruner DW, Movsas B. Recommended patient-reported core set of symptoms to measure in head and neck cancer treatment trials. J Natl Cancer Inst. 2014;106(7). PubMed PMID: 25006189.

Gebre-Medhin M, Brun E, Engström P, Haugen Cange H, Hammarstedt-Nordenvall L, Reizenstein J, Nyman J, Abel E, Friesland S, Sjödin H, Carlsson H, Söderkvist K, Thomasson M, Zackrisson B, Nilsson P. ARTSCAN III: A Randomized Phase III Study Comparing Chemoradiotherapy With Cisplatin Versus Cetuximab in Patients With Locoregionally Advanced Head and Neck Squamous Cell Cancer. J Clin Oncol. 2021 Jan 1;39(1):38-47. doi: 10.1200/JCO.20.02072. Epub 2020 Oct 14. PMID: 33052757; PMCID: PMC7771720.

Gillison ML, Trotti AM, Harris J, et al. Radiotherapy plus cetuximab or cisplatin in human papillomavirus-positive oropharyngeal cancer (NRG Oncology RTOG 1016): a randomised, multicentre, non-inferiority trial. Lancet. 2019;393(10166):40-50. doi: 10.1016/S0140-6736(18)32779-X. PubMed PMID: 30449625; PubMed Central PMCID: PMC6541928.

NVMO-commissie ter Beoordeling van Oncologische Middelen (BOM). PASKWIL criteria 2018.

Mehanna H, Robinson M, Hartley A, et al. Radiotherapy plus cisplatin or cetuximab in low-risk human papillomavirus-positive oropharyngeal cancer (De-ESCALaTE HPV): an open-label randomised controlled phase 3 trial. Lancet. 2019;393(10166):51-60. PubMed PMID: 30449623; PubMed Central PMCID: PMC6319250.

Windon MJ, Fakhry C, Faraji F, et al. Priorities of human papillomavirus-associated oropharyngeal cancer patients at diagnosis and after treatment. Oral Oncol. 2019;95:11-15. PubMed PMID: 31345377; PubMed Central PMCID: PMC6662631.

Yom SS, Torres-Saavedra P, Caudell JJ, Waldron JN, Gillison ML, Xia P, Truong MT, Kong C, Jordan R, Subramaniam RM, Yao M, Chung CH, Geiger JL, Chan JW, O'Sullivan B, Blakaj DM, Mell LK, Thorstad WL, Jones CU, Banerjee RN, Lominska C, Le QT. Reduced-Dose Radiation Therapy for HPV-Associated Oropharyngeal Carcinoma (NRG Oncology

HN002). J Clin Oncol. 2021 Mar 20;39(9):956-965. doi: 10.1200/JCO.20.03128. Epub 2021 Jan 28. PMID: 33507809.

Geldigheid en Onderhoud

Module[1]	Regiehouder(s)[2]	Jaar van autorisatie	Eerstvolgende beoordeling actualiteit richtlijn[3]	Frequentie van beoordeling op actualiteit[4]	Wie houdt er toezicht op actualiteit[5]	Relevante factoren voor wijzigingen in aanbeveling[6]

[1] Naam van de module
[2] Regiehouder van de module (deze kan verschillen per module en kan ook verdeeld zijn over meerdere regiehouders)
[3] Maximaal na vijf jaar
[4] (half)Jaarlijks, eens in twee jaar, eens in vijf jaar
[5] regievoerende vereniging, gedeelde regievoerende verenigingen, of (multidisciplinaire) werkgroep die in stand blijft
[6] Lopend onderzoek, wijzigingen in vergoeding/organisatie, beschikbaarheid nieuwe middelen

Bijlagen bij module 12.4

Kennislacunes

What are the effects of a de-escalating strategy in the treatment of human papilloma virus (HPV)-associated oropharyngeal cancer (OPC) be used compared to care-as-usual for OPC?

Implementatieplan

Aanbeveling	Tijdspad voor implementatie: < 1 jaar, 1 tot 3 jaar of 3 tot 5 jaar	Verwacht effect op kosten	Randvoorwaarden voor implementatie (binnen aangegeven tijdspad)	Mogelijke barrières voor implementatie[1]	Te ondernemen acties voor implementatie[2]	Verantwoordelijken voor acties[3]	Overige opmerkingen
De-escaleer de behandeling van patiënten met een HPV-positief orofarynxcarcinoom alleen in studieverband.	<1 jaar	Geen	Geen	Er kunnen mogelijk geen lopende studies zijn die de-escalatie aanbieden.	Indien er geen lopende studies zijn die de-escalatie aanbieden zou de kennislacune opgenomen kunnen worden in een kennisagenda van een relevante wetenschappelijke vereniging.	NIV / NVRO (besluit opname kennisagenda)	Geen

[1] Barrières kunnen zich bevinden op het niveau van de professional, op het niveau van de organisatie (het ziekenhuis) of op het niveau van het systeem (buiten het ziekenhuis). Denk bijvoorbeeld aan onenigheid in het land met betrekking tot de aanbeveling, onvoldoende motivatie of kennis bij de specialist, onvoldoende faciliteiten of personeel, nodige concentratie van zorg, kosten, slechte samenwerking tussen disciplines, nodige taakherschikking, et cetera.

[2] Denk aan acties die noodzakelijk zijn voor implementatie, maar ook acties die mogelijk zijn om de implementatie te bevorderen. Denk bijvoorbeeld aan controleren aanbeveling tijdens kwaliteitsvisitatie, publicatie van de richtlijn, ontwikkelen van implementatietools, informeren van ziekenhuisbestuurders, regelen van goede vergoeding voor een bepaald type behandeling, maken van samenwerkingsafspraken.

[3] Wie de verantwoordelijkheden draagt voor implementatie van de aanbevelingen, zal tevens afhankelijk zijn van het niveau waarop zich barrières bevinden. Barrières op het niveau van de professional zullen vaak opgelost moeten worden door de beroepsvereniging. Barrières op het niveau van de organisatie zullen vaak onder verantwoordelijkheid van de ziekenhuisbestuurders vallen. Bij het oplossen van barrières op het niveau van het systeem zijn ook andere partijen, zoals de NZA en zorgverzekeraars, van belang.

Evidence tables

Evidence table for intervention studies (randomized controlled trials)

Research question: Should a de-escalating strategy in the treatment of human papilloma virus (HPV)-associated oropharyngeal cancer (OPC) be used compared to care-as-usual for OPC?

Study reference	Study characteristics	Patient characteristics	Intervention (I)	Comparison / control (C)	Follow-up	Outcome measures and effect size	Comments
Gillison 2019 (RTOG 1016)	Type of study: RCT (a randomised, multicentre, non-inferiority trial) Setting and country: health-care centres in the USA and Canada Funding and conflicts of interest: National Cancer Institute USA, Eli Lilly, and The Oral Cancer Foundation *"The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding authors had full access to all the data in the study and had final responsibility for the decision to submit for publication."*	Inclusion criteria: histologically confirmed HPV-positive oropharyngeal carcinoma; American Joint Committee on Cancer 7th edition9 clinical categories T1–T2, N2a–N3 M0 or T3–T4, N0–N3 M0; Zubrod performance status 0 or 1; age at least 18 years; and adequate bone marrow, hepatic, and renal function. Exclusion criteria: see online appendix N total at baseline: Intervention: 399 Control: 406 Groups comparable at baseline. Patients were predominantly men and white and had a median age of 58 years (IQR 52–63).	Intravenous **cetuximab** (Eli Lilly; Indianapolis, IN, USA) at a loading dose of 400 mg/m² 5–7 days before radiotherapy initiation, followed by cetuximab 250 mg/m² weekly for seven doses (total 2150 mg/m²) All patients received accelerated intensity-modulated radiotherapy delivered at 70 Gy in 35 fractions over 6 weeks at six fractions per week (with two fractions given on one day, at least 6 h apart).	**Cisplatin** (commercially available and obtained by each individual institution) 100 mg/m² on days 1 and 22 of radiotherapy (total 200 mg/m²) All patients received accelerated intensity-modulated radiotherapy delivered at 70 Gy in 35 fractions over 6 weeks at six fractions per week (with two fractions given on one day, at least 6 h apart).	Length of follow-up: Median: 4.5 years Loss-to-follow-up: Intervention: 23 (6%) Reasons: 22 patient withdrew consent, 1 unknown reason Further: 19 discontinued intensity-modulated radiotherapy, 66 discontinued cetuximab) Control: 31 (8%) Reasons: 26 patient withdrew consent, 5	Outcome measures and effect size: Mortality (n): I: 78 (59%) C: 55 (41%) 5-year overall survival: I: 77.9% (95% CI 73.4–82.5) C: 84.6% (95% CI 80.6–88.6) HR: 1.45, one-sided 95% upper CI 1.94; p=0.5056 Progression: I: 122 (62%) C: 76 (38%) 5-year progression free survival: I: 67.3% (95% CI 62.4–72.2) C: 78.4% (95% CI 73.8–83.0) Complications/adverse events Acute period patient total	

					unknown reason Further: 14 discontinued intensity-modulated radiotherapy, 18 discontinued cisplatin) <u>Incomplete outcome data</u>: Not reported	I: 394 (99%) C: 398 (98%) Late period patient total I: 375 (94%) C: 383 (94%) <u>Quality of life</u> *To be reported.* <u>Functional outcomes</u> EORTC QLQ-H&N35 (Patient-reported severity of swallowing problems) Mean change score from baseline to end of treatment: I: 47.4 (n=132 in analysis) C: 48.0 (n=131 in analysis); p=0·86 At 1-year follow-up: I: 7.6 (n=116 in analysis) C: 2.5 (n=119 in analysis) Dental health 1 year after treatment (normal or mild changes or good dental health) I: 223/267 (84%), mean 1.64 teeth lost. C: 233/367 (87%), mean 1.05 teeth lost. <u>(Work) Participation</u>	

						Not reported	
Mehanna 2019 (De-ESCALaTE HPV)	Type of study: RCT (a randomised, multicentre trial) Setting and country: health-care centres in Ireland, the Netherlands, and the UK. Funding and conflicts of interest: Cancer Research UK	Inclusion criteria: aged at least 18 years with a histologically confirmed diagnosis of advanced oropharyngeal squamous cell carcinoma tumour, node, and metastasis (TNM) 7th Edition. manual: T3N0–T4N0, and T1N1–T4N3) that was classified as low risk as per the Ang classification:4 that is, the tumour sample had to be positive on p16 immunohistochemistry, and the patient had to be a non-smoker or have a lifetime self-reported smoking history of less than 10 pack-years. Patients had to have an Eastern Cooperative Oncology Group (ECOG) performance status score of 0 or 1, and adequate renal, haematological, and hepatic function for cisplatin-based curative chemoradiotherapy. Exclusion criteria: T1–T2N0 disease or were classified as HPV-negative, high-risk, or HPV-positive oropharyngeal squamous cell carcinoma intermediate risk on the Ang classification. N total at baseline:	Intravenous **cetuximab** 400 mg/m² loading dose 1 week before followed by seven weekly infusions of 250 mg/m² during radiotherapy. All patients received accelerated intensity-modulated radiotherapy delivered at 70 Gy in 35 fractions over 7 weeks at five fractions per week.	Three doses of intravenous **cisplatin** 100 mg/m² on days 1, 22, and 43 of radiotherapy. All patients received accelerated intensity-modulated radiotherapy delivered at 70 Gy in 35 fractions over 7 weeks at five fractions per week.	Length of follow-up: Median: 25.9 months (95% CI 25.5–26.0) Loss-to-follow-up: Intervention: 26 (15%) Reasons: 7 withdrawals, 19 death Control: 13 (8%) Reasons: 7 withdrawals, 6 death. Number of patients in analysis primary outcome: Intervention: 165 Control: 162	Outcome measures and effect size (include 95%CI and p-value if available): Mortality (n): I: 20 (11.9%) C: 6 (3.6 %) 2-year overall survival: I: 89.4% C: 97.5% HR: 5.0 (95% CI 1.7–14.7) Progression: Not reported 2-year progression free survival: Not reported. Complications/adverse events Overall severe (grades 3–5) toxicity (mean number of events per patient): I: 4.8 (4·2–5·4) C: 4.8 (95% CI 4.2–5.4); p=0.98 Acute period all-grade toxicity (mean number of events per patient)	

		Intervention: 168 Control: 166 Groups comparable at baseline.				I: 20.4 (95% CI 19.2–21.5) C: 20.0 (95% CI 18.8–21.1); p=0.64 Late period all-grade toxicity (mean number of events per patient) I: 9.9 C: 9.4; p=0.49 <u>Quality of life (EORTC)</u> Physical functioning: n.s. Role functioning: n.s. Emotional functioning: n.s. Cognitive functioning: n.s. Social functioning: n.s. <u>Functional outcomes</u> Swallowing (Global M.D. Anderson Dysphagia Inventory score) Mean difference at 2-year follow-up: MD: 6.90 in favour of cisplatin; p=0.1279 <u>(Work) Participation</u> Not reported	

Risk of bias table for intervention studies (randomized controlled trials)

Research question: Should a de-escalating strategy in the treatment of human papilloma virus (HPV)-associated oropharyngeal cancer (OPC) be used compared to care-as-usual for OPC?

Study reference (first author, publication year)	**Describe method of randomisation[1]**	**Bias due to inadequate concealment of allocation?[2]** (unlikely/likely/unclear)	**Bias due to inadequate blinding of participants to treatment allocation?[3]** (unlikely/likely/unclear)	**Bias due to inadequate blinding of care providers to treatment allocation?[3]** (unlikely/likely/unclear)	**Bias due to inadequate blinding of outcome assessors to treatment allocation?[3]** (unlikely/likely/unclear)	**Bias due to selective outcome reporting on basis of the results?[4]** (unlikely/likely/unclear)	**Bias due to loss to follow-up?[5]** (unlikely/likely/unclear)	**Bias due to violation of intention to treat analysis?[6]** (unlikely/likely/unclear)
Gillison 2019	Randomly assigned patients (1:1) to receive either radiotherapy plus cetuximab or radiotherapy plus cisplatin. Randomisation was balanced by using randomly permuted blocks, and patients were stratified by T category (T1–T2 versus T3–T4), N category (N0–N2a versus N2b–N3), Zubrod performance status (0 versus 1), and tobacco smoking history (≤10 pack-years versus >10 pack-years). Treatment assignment was centrally generated at the NRG Oncology	Unlikely	Unclear, in the article is only stated "Treatment assignment was not masked to the participating site, the enrolling physician, or the responsible statistician."	Likely, it is assumed that the care provider is the enrolling physician.	Unclear, in the article is only stated "Treatment assignment was not masked to the participating site, the enrolling physician, or the responsible statistician."	Unlikely.	Unlikely. Intervention: 23 (6%) lost to follow-up Control: 31 (8%)	Unlikely. "We based our primary analysis on the modified intention-to-treat approach, whereby all patients meeting eligibility criteria are included."

	Statistics and Data Management Center (Philadelphia, PA, USA) and provided to the institution when the patient was entered.							
Mehanna 2019	Eligible patients underwent computer-generated central randomisation. Patients were randomly assigned in a 1:1 ratio to receive cisplatin-based chemoradiotherapy or cetuximab bioradiotherapy. Trial-group assignments were balanced by use of a bespoke minimisation algorithm according to centre, tumour stage (TNM7: T1–T2 versus T3–T4), nodal stage (N0–1 versus N2–3), radiotherapy site (unilateral; bilateral), and planned gastrostomy insertion before treatment.	Unlikely	Likely. Open-label trial.	Likely Open-label trial.	Unclear	Unlikely.	Unlikely Intervention: 26 (15%) lost to follow-up (19 death) Control: 13 (8%)	Unlikely. “An intention-to-treat analysis was done for all outcomes, and a per-protocol analysis was done for primary outcomes and secondary outcomes of toxicity and survival.”

Table of excluded studies

Author and year	Reason for exclusion
Chera 2019	Not a comparative study, but a single-arm trial of de-intensified chemoradiotherapy.
Rühle 2019	Article in German
Seiwert 2019	Not a comparative study, but a single-arm trial of dose and volume de-escalated radiotherapy or chemoradiotherapy based on response to induction chemotherapy.
Bahig 2018	Not a comparative study, but a single-arm, phase 2 trial of de-escalated chemoradiation therapy for HPV-positive oropharyngeal carcinoma.
Hegde 2018	Not a comparative study, but a single-arm, phase 2 trial of de-escalated chemoradiation therapy for HPV-positive oropharyngeal carcinoma.
Hegde 2018	Not a comparative study, but a single-arm, phase 2 trial of de-escalated chemoradiation therapy for HPV-positive oropharyngeal carcinoma.
Lassen 2018	Not a comparative study, but a prognotic factors study.
Pearlstein 2018	Not a comparative study, but describing outcomes from 2 prospective studies examining de-intensified chemoradiotherapy.
Sinha 2018	Intervention and control do not match the PICO: primary surgical versus non-surgical treatment.
Stock 2018	Not a systematic review, but a narrative review
Buglione 2017	Included in Suton 2019, not a randomized trial but a subgroup analysis of a phase II trial (n=30)
De Santis 2017	Not a comparative study, but a descriptive study of characteristics of oropharyngeal human papillomavirus-associated cancers.
Tassone 2017	Not a comparative study, but a prognotic factors study.
Winquist 2017	Patient population does not match the PICO: stratification of outcomes by HPV-status was not available.
Winquist 2017	Patient population does not match the PICO: stratification of outcomes by HPV-status was not available.
Argiris 2016	Patient population does not match the PICO: stratification of outcomes by HPV-status was not available.
Marcu 2016	Not a systematic review, but a narrative review.
Huang 2016	More recent SR addressing the same PICO available (i.e. Suton 2019).
Naghavi 2016	Not a systematic review, but a narrative review.
Shaikh 2015	Not a comparative study, but a descriptive study.
Wang 2015	Intervention and control do not match the PICO: primarily surgery versus primarily radiation.
Mak 2014	Patient population does not match PICO: locally advanced head and neck squamous cell carcinoma, not stratified by HPV-status.
O'Leary 2014	Intervention does not match PICO: study is about transoral robotic surgery.
Petrelli 2014	Not a comparative study, but a prognotic factors study.

Table of ongoing RCTs.

1. Cetuximab-based bioradiotherapy versus cisplatin
TROG12.01 trial (NCT01855451) will evaluate in patients with stage III/IV HPV-positive OPC randomized to receive cetuximab versus weekly cisplatin with concurrent standard dose radiation therapy.
2. Reduced-intensity chemoradiation or omission of chemo
Trial NCT02254278 will evaluate reduced-dose IMRT in nonsmoking p16-positive LA-OPC patients. Patients in one arm will receive 60 Gy IMRT over 5 weeks along with weekly cisplatin, and in the other arm, 60 Gy IMRT alone over 6 weeks. [werd na de zoekdatum gepubliceerd en werd daarom besproken in de overwegingen]
4. Minimally invasive transoral surgery followed by de-intensified adjuvant therapy (either omission of chemotherapy or reduced-dose radiotherapy)
ADEPT (NCT01687413) is a phase III trial comparing postoperative radiotherapy with or without cisplatin in HPV-positive T1-4a OPSCC patients. Included patients must have received minimally invasive surgery and demonstrated extra-capsular spread from disease in the neck.
ECOG-E3311 (NCT01898494) is a phase II trial of treatment for HPV-positive locally advanced OPSCC (stages III-IVa + IVb without distant metastasis). Patients are stratified after minimally invasive surgery. Medium-risk patients are randomised to either standard or reduced-dose radiotherapy.
PATHOS (NCT02215265) is a phase III trial of treatment for HPV-positive OPSCC (T1-3, N0-2b). Patients are stratified after minimally invasive surgery. Medium-risk patients are randomised to either standard or

reduced-dose radiotherapy. High-risk patients are randomised to radiotherapy with or without concurrent cisplatin.
SIRS trial (NCT02072148) is a trial in HPV-positive OPCs that evaluates adjuvant treatment stratified by risk based on histopathologic features after minimally invasive surgery. The low risk group will be followed with PET-computed tomography (PET-CT). Intermediate risk and high-risk groups will receive adjuvant radiation therapy or concurrent chemoradiotherapy (CRT), respectively

Literature search strategy

PubMed November 2019

Search	Query	Items found
#44	Search #42 AND #13	217
#43	Search #42 AND #12	23
#42	Search #39 OR #41	705
#41	Search #28 AND #38 AND #40	168
#40	Search de escalat*(tiab) OR de intensif*(tiab)	2129
#39	Search #25 AND #28 AND #38	642
#38	Search "Oropharyngeal Neoplasms"(Mesh) OR "oropharyngeal tumor"(tiab) OR "oropharyngeal tumour"(tiab) OR "oropharyngeal tumours"(tiab) OR "oropharynx tumors"(tiab) OR "oropharyngeal cancer"(tiab) OR "oropharyngeal cancers"(tiab) OR "oropharynx cancer"(tiab) OR "oropharynx cancers"(tiab) OR "oropharyngeal carcinoma"(tiab) OR "oropharyngeal carcinomas"(tiab) OR "oropharynx carcinoma"(tiab) OR "oropharynx carcinomas"(tiab) OR "oropharyngeal neoplasm"(tiab) OR "oropharyngeal neoplasms"(tiab) OR "oropharynx neoplasm"(tiab) OR "oropharynx neoplasms"(tiab)	9986
#28	Search "Papillomaviridae"(Mesh) OR human papilloma virus(tiab) OR wart virus(tiab) OR condyloma virus(tiab) OR hpv(tiab) OR human papillomavirus(tiab) OR verruca virus(tiab) OR viral verruca(tiab) OR virus verruca(tiab) OR virus wart(tiab)	54437
#25	Search "Radiotherapy"(Mesh) OR "radiotherapy"(Subheading) OR "Radiosurgery"(Mesh) OR radiotherap*(tiab) OR radiati*(tiab) OR radiosurg*(tiab) OR irradiati*(tiab) OR "x ray therapy" (tiab) OR "x ray therapies" (tiab) OR radioimmunotherap*(tiab) OR immunoradiotherap*(tiab)	707945
#17	Search #11 AND #14	41
#16	Search #11 AND #13	12
#15	Search #11 AND #12	2
#14	Search "Epidemiologic Studies"(Mesh) OR cohort(tiab) OR (case(tiab) AND (control(tiab) OR controll*(tiab) OR comparison(tiab) OR referent(tiab))) OR risk(tiab) OR causation(tiab) OR causal(tiab) OR "odds ratio"(tiab) OR etiol*(tiab) OR aetiol*(tiab) OR "natural history"(tiab) OR predict*(tiab) OR prognos*(tiab) OR outcome(tiab) OR course(tiab) OR retrospect*(tiab)	6369580
#13	Search randomized controlled trial(pt) OR controlled clinical trial(pt) OR randomized(tiab) OR placebo(tiab) OR drug therapy(sh) OR randomly(tiab) OR trial(tiab) OR groups(tiab)	4621104
#12	Search ("Meta-Analysis as Topic"(Mesh) OR “Meta-Analysis”(Publication Type) OR metaanaly*(tiab) OR metanaly*(tiab) OR meta-analy*(tiab) OR meta synthes*(tiab) OR metasynthes*(tiab) OR meta ethnograph*(tiab) OR metaethnograph*(tiab) OR meta summar*(tiab) OR metasummar*(tiab) OR meta-aggregation(tiab) OR metareview(tiab) OR meta-review(tiab) OR overview of reviews(tiab) OR ((systematic*(ti) OR scoping(ti) OR umbrella(ti) OR meta-narrative(ti) OR metanarrative(ti) OR evidence based(ti)) AND (review*(ti) OR overview*(ti))) OR ((evidence(ti) OR narrative(ti) OR metanarrative(ti) OR qualitative(ti)) AND synthesis(ti)) OR systematic review(pt) OR prisma(tiab) OR preferred reporting items(tiab) OR quadas*(tiab) OR systematic review*(tiab) OR systematic literature(tiab) OR structured literature search(tiab) OR systematic overview*(tiab) OR scoping review*(tiab) OR umbrella review*(tiab) OR mapping review*(tiab) OR systematic mapping(tiab) OR evidence synthes*(tiab) OR narrative synthesis(tiab) OR metanarrative synthesis(tiab) OR research synthesis(tiab) OR qualitative synthesis(tiab) OR realist	301245

	synthesis(tiab) OR realist review(tiab) OR realist evaluation(tiab) OR systematic qualitative review(tiab) OR mixed studies review(tiab) OR mixed methods synthesis(tiab) OR mixed research synthesis(tiab) OR quantitative literature review(tiab) OR systematic evidence review(tiab) OR evidence-based review(tiab) OR comprehensive literature search(tiab) OR integrated review*(tiab) OR integrated literature review(tiab) OR integrative review*(tiab) OR integrative literature review*(tiab) OR structured literature review*(tiab) OR systematic search and review(tiab) OR meta-narrative review*(tiab) OR metanarrative review(tiab) OR systematic narrative review(tiab) OR systemic review(tiab) OR systematized review(tiab) OR systematic research synthesis(tiab) OR bibliographic*(tiab) OR hand-search*(tiab) OR handsearch*(tiab) OR manual search*(tiab) OR searched manually(tiab) OR manually searched(tiab) OR journal database*(tiab) OR review authors independently(tiab) OR reviewers independently(tiab) OR independent reviewers(tiab) OR independent review authors(tiab) OR electronic database search*(tiab) OR (study selection(tiab) AND data extraction(tiab)) OR (selection criteria(tiab) AND data collection(tiab)) OR (selection criteria(tiab) AND data analysis(tiab)) OR (evidence acquisition(tiab) AND evidence synthesis(tiab))) NOT ("Comment" (Publication Type) OR "Letter" (Publication Type)) NOT (“Animals”(Mesh) NOT “Humans”(Mesh))	

Embase November 2019

No.	Query	Results
#19	#14 AND (#16 OR #17)	**1**
#18	#13 AND (#16 OR #17)	**1**
#17	#9 AND #15 NOT ‘conference abstract’:it	**258**
#16	#8 AND #15 NOT ‘conference abstract’:it	**41**
#15	#11 OR #12	**1856**
#14	30449625	**2**
#13	30887169	**1**
#12	#3 AND #7	**1788**
#11	#3 AND #10	**358**
#10	'de escalat*':ti,ab,kw OR 'de intensif*':ti,ab,kw	**4419**
#9	('clinical trial'/exp OR 'randomization'/exp OR 'single blind procedure'/exp OR 'double blind procedure'/exp OR 'crossover procedure'/exp OR 'placebo'/exp OR 'prospective study'/exp OR rct:ab,ti OR random*:ab,ti OR 'single blind':ab,ti OR 'randomised controlled trial':ab,ti OR 'randomized controlled trial'/exp OR placebo*:ab,ti) NOT 'conference abstract':it	**2301957**
#8	('meta analysis'/de OR cochrane:ab OR embase:ab OR psycinfo:ab OR cinahl:ab OR medline:ab OR ((systematic NEAR/1 (review OR overview)):ab,ti) OR ((meta NEAR/1 analy*):ab,ti) OR metaanalys*:ab,ti OR 'data extraction':ab OR cochrane:jt OR 'systematic review'/de) NOT (('animal experiment'/exp OR 'animal model'/exp OR 'nonhuman'/exp) NOT 'human'/exp)	**459329**
#7	#4 OR #5 OR #6	**636719**
#6	'cancer radiotherapy'/exp OR 'cancer radiation':ti,ab OR 'cancer radiotherapy':ti,ab OR 'tumor irradiation':ti,ab OR 'tumor radiation':ti,ab OR 'tumor radiotherapy':ti,ab OR 'tumour irradiation':ti,ab OR 'tumour radiation':ti,ab OR 'tumour radiotherapy':ti,ab	**233749**
#5	'chemoradiotherapy'/de OR chemoradiotherap*:ti,ab	**54192**
#4	'radiotherapy'/exp OR 'bioradiant therapy':ti,ab OR 'bucky irradiation':ti,ab OR 'bucky radiation':ti,ab OR 'bucky radiotherapy':ti,ab OR 'bucky ray':ti,ab OR 'bucky therap*':ti,ab	**633751**

	OR 'hemibody irradiation':ti,ab OR 'hypophysis irradiation':ti,ab OR 'hypophysis radiation':ti,ab OR 'irradiation therapy':ti,ab OR 'irradiation treatment':ti,ab OR 'lymphatic irradiation':ti,ab OR 'pituitary irradiation':ti,ab OR 'radiation beam centration':ti,ab OR 'radiation repair':ti,ab OR 'radiation therapy':ti,ab OR 'radiation treatment':ti,ab OR 'radio therapy':ti,ab OR 'radio treatment':ti,ab OR 'radiohypophysectomy':ti,ab OR 'radiotherapy':ti,ab OR 'radiotreatment':ti,ab OR 'roentgen therapy':ti,ab OR 'roentgen treatment':ti,ab OR 'rontgen therapy':ti,ab OR 'therapeutic radiology':ti,ab OR 'x radiotherapy':ti,ab OR 'x ray therapy':ti,ab OR 'x ray treatment':ti,ab OR 'x-ray therapy':ti,ab	
#3	#1 AND #2	**5061**
#2	'wart virus'/exp OR 'human papilloma virus':ti,ab OR 'wart virus':ti,ab OR 'condyloma virus':ti,ab OR 'hpv':ti,ab OR 'human papillomavirus':ti,ab OR 'verruca virus':ti,ab OR 'viral verruca':ti,ab OR 'virus verruca':ti,ab OR 'virus wart':ti,ab OR 'verruca, viral':ti,ab OR 'papillomaviridae'/exp	**75826**
#1	'oropharynx cancer'/exp OR ((oropharyn* NEAR/3 (tumor* OR tumour* OR cancer* OR carcinoma* OR neoplasm*)):ti,ab,kw)	**14431**

Bijlage 1 Kennislacunes

Inleiding
Tijdens de herziening van de richtlijn 'Hoofd-Halstumoren' is systematisch gezocht naar onderzoeksbevindingen die nuttig konden zijn voor het beantwoorden van de uitgangsvragen. Een deel (of een onderdeel) van de hiervoor opgestelde zoekvragen is met het resultaat van deze zoekacties te beantwoorden, een groot deel echter niet. Door gebruik te maken van de evidence-based methodiek (EBRO) is duidelijk geworden dat er nog kennislacunes bestaan. De werkgroep is van mening dat (vervolg)onderzoek wenselijk is om in de toekomst een duidelijker antwoord te kunnen geven op vragen uit de praktijk. Om deze reden heeft de werkgroep per module aangegeven op welke vlakken nader onderzoek gewenst is.

Moduletitel	Kennislacune
Diagnostiek invasiediepte mondholtecarcinoom	What is the agreement between preoperative clinical examination (by palpation), computed tomography (CT), positron emission tomography/computed tomography (PET-CT), magnetic resonance imaging (MRI) or intraoral ultrasound, and postoperative histopathologic results for measuring the depth of the invasion by a tumor in patients with an oral cavity carcinoma?
Diagnostiek botinvasie mandibula	What is the diagnostic accuracy of SPECT, DECT, CBCT, CT, PET-CT, MRI, OPG, or diagnostic algorithms (with a pathology assessment as a reference standard) to diagnose mandibular or bone marrow invasion of a tumor in patients with an oral cavity carcinoma?
HPV-statusbepaling	What is the diagnostic accuracy of diagnostic test algorithms to determine the HPV-status on histological material in patients with an oropharyngeal carcinoma? What is the diagnostic accuracy of tests on cytologic material to determine the HPV-status in patients with a carcinoma of unknown primary?
Aanvraag en verslag pathologie-onderzoek	Geen
Behandeling T1-2N0-1 orofarynxcarcinomen	What are the (un)beneficial effects of TORS (with or without microsurgery or adjuvant radiotherapy) on overall survival, disease-free survival, local control, swallowing, trismus, taste, dryness of the mouth, mucositis, and quality of life in patients with a T1-2N0-1 oropharyngeal carcinoma, when compared to primary radiotherapy.
Behandeling HPV-positieve tumoren	What are the effects of a de-escalating strategy in the treatment of human papilloma virus (HPV)-associated oropharyngeal cancer (OPC) be used compared to care-as-usual for OPC?