Review

Intravoxel incoherent motion magnetic resonance imaging in head and neck cancer: A systematic review of the diagnostic and prognostic value

Daniel P. Noij [a,*], Roland M. Martens [a], J. Tim Marcus [b], Remco de Bree [c], C. René Leemans [d], Jonas A. Castelijns [a], Marcus C. de Jong [a], Pim de Graaf [a]

[a] Department of Radiology and Nuclear Medicine, VU University Medical Center, De Boelelaan 1117, PO Box 7057, 1007 MB Amsterdam, The Netherlands
[b] Department of Physics and Medical Technology, VU University Medical Center, De Boelelaan 1117, PO Box 7057, 1007 MB Amsterdam, The Netherlands
[c] Department Head and Neck Surgical Oncology, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands
[d] Department of Otolaryngology – Head and Neck Surgery, VU University Medical Center, De Boelelaan 1117, PO Box 7057, 1007 MB Amsterdam, The Netherlands

ARTICLE INFO

ABSTRACT

Intravoxel incoherent motion (IVIM) imaging is increasingly applied in the assessment of head and neck cancer (HNC). Our purpose was to determine the diagnostic and prognostic performance of IVIM in HNC by performing a critical review of the literature. Pubmed and EMBASE were searched until May 2016. Study and patients characteristics, imaging protocol and diagnostic or prognostic outcomes were extracted by 2 independent reviewers. The studied IVIM parameters were diffusion coefficient (D), pseudodiffusion coefficient (D∗), and perfusion fraction (f). We included 10 diagnostic studies, 5 prognostic studies and 2 studies assessing both. Studies were very heterogeneous in terms of applied b-values, imaging protocols, outcome measurements and reference standards; therefore we did not perform a meta-analysis. The most commonly used sequence was "spin-echo planar imaging". A median of 10.5 b-values (range, 3–17) were used. All but three studies included at least 4 b-values below b = 200 s/mm^2. By combining IVIM-parameters squamous cell carcinomas, lymphomas, malignant salivary gland tumors, Warthin's tumors and pleomorphic adenomas could be differentiated with a sensitivity of 85–87% and specificity of 80–100%. Low pre-treatment D or f and an increase in D during treatment were associated with a favorable response to treatment. D∗ appeared to be the parameter with the lowest prognostic value. Future research should focus on finding the optimal IVIM protocol, using uniformly accepted study methods and larger patient populations.

Introduction

Head and neck cancer (HNC) accounts for approximately 4% of the cancer case worldwide, making HNC the sixth most common cancer by incidence rate [1,2]. HNC mainly consists of tumors arising in the oral cavity, nasopharynx, oropharynx, hypopharynx, larynx and salivary glands.

* Corresponding author.
E-mail address: d.noij@vumc.nl (D.P. Noij).

Squamous cell carcinomas (SCC) account for over 90% of HNC [3]. Alcohol and tobacco use are the most important risk factors [4]. While early stage disease is usually treated by surgery or radiotherapy, advanced stage disease is generally treated by surgery and adjuvant radiotherapy with or without chemotherapy or combined chemotherapy and radiotherapy. Salvage surgery is then held in reserve for residual or recurrent disease [2,5–7]. While chemotherapy is mainly used in a concomitant setting with radiotherapy, in selected cases it can also be applied as neoadjuvant treatment [2]. There is increasing evidence that in some geographic regions up to 80% of the oropharyngeal SCC is associated with the human papillomavirus (HPV), especially in relatively young patients who do not drink or smoke [8]. HPV-associated oropharyngeal SCC has a different tumor biology and is associated with a better prognosis than HPV-negative SCC [5–7]. Therefore it is proposed to de-escalate treatment in HPV-associated oropharyngeal SCC in patients who do not smoke.

Nasopharyngeal carcinoma (NPC) takes a unique place in epithelial HNC because of the very distinct geographical distribution ranging from 1:100.000 in Western Europe to >20:100.000 in parts of Southeast Asia [2,9]. Further it harbors an association with the Epstein-Barr virus (EBV) which is not seen in other HNC [9]. NPC has a different tumor biology as compared to other HNC.

Imaging is increasingly used for diagnosing and staging of HNC, monitoring the effect of treatment and in the detection of distant metastases and recurrent disease [10–12]. In this systematic review we focus on the use of intravoxel incoherent motion (IVIM) magnetic resonance imaging (MRI) for diagnosis in HNC.

In general, water diffusion is restricted in malignant tissue. With diffusion-weighted imaging (DWI) this restricted diffusion can be imaged and quantified. The main advantage of DWI compared to other functional imaging techniques (e.g. dynamic contrast-enhanced MRI and positron-emission tomography) is that it requires neither the administration of contrast medium or radioactive tracer nor the use of ionizing radiation.

One of the proposed methods to quantify diffusion is by considering diffusion as a mono-exponential phenomenon. In this way diffusion can be quantified in an apparent diffusion coefficient (ADC) [13]. The word "apparent" implies that in this way true diffusion is not measured. Especially at low b-values other parameters as blood volume and blood flow also contribute to the ADC [14,15]. The ADC-concept provides a quantifiable measure with promising results in HNC, e.g. in discriminating metastatic from benign lymph nodes with an accuracy of >85% and in the detection of recurrent disease with an accuracy of >78% [16].

The signal decay after the diffusion-encoding gradients is not only caused by diffusion, but also by pseudorandom, or "incoherent", perfusion at the capillary level. To account for this, Le Bihan et al. introduced the bi-exponential IVIM model [13,1]:

$$\frac{S_b}{S_0} = (1-f) \cdot e^{(-bD)} + f \cdot e^{(-bD*)}$$

where S_b represents the signal intensity with diffusion gradient b, and S_0 represents the signal intensity without diffusion gradients. D is known as pure or slow diffusion coefficient which is related to pure molecular diffusion. $D*$ is the fast or pseudodiffusion coefficient that resembles the perfusion related incoherent microcirculation and is about a factor of 10 greater than D in biological tissue [13]. Finally, f is the perfusion or (micro) vascular volume fraction which depends on capillary geometry and blood velocity [13]. In this way pure tissue diffusion may be quantified and also perfusion characteristics may be assessed without the admission of contrast-material. Commonly D is first estimated using a linear fit using only high b-values (i.e., above 200 s/mm^2 [17]) and then f and $D*$ are calculated using a non-linear least-squares algorithm.

With IVIM being increasingly used in HNC, a critical systematic review of the diagnostic and prognostic value of this technique is warranted. The purpose of this study was therefore to determine the diagnostic and prognostic performance of IVIM in HNC. Histopathology, other imaging modalities or clinical follow-up were used as reference standards.

Methods and materials

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement for systematic reviews and meta-analyses was used as a guidance [18].

Search strategy

This systematic search was conducted in Pubmed and Embase until May 2016 for original articles on the diagnostic and/or prognostic capability of IVIM in HNC. We did not apply language restrictions. We approached corresponding authors for additional data if necessary (e.g. to compute sensitivity and specificity). The only included search terms were "(IVIM OR ((intra-voxel OR intravoxel) AND incoherent AND motion))" in order to be as sensitive as possible. In the Pubmed search we used text words [tw] in the absence of MeSH-terms on this subject.

Two authors (D.P.N. and R.M.M.) independently selected relevant articles based on title and abstracts and discrepancies were resolved by consensus.

The inclusion criteria were: (1) The study population consisted of at least 10 patients with malignant lesions in the head and neck area; (2) The study assessed diagnosing malignancy, response prediction to therapy, detection of residual/recurrent disease. Or data of these subjects could be extracted from the article; (3) Histopathology, clinical follow-up or another imaging modality was used as reference standard test.

Exclusion criteria were: (1) The publication was a review, meta-analysis, only published as abstract or if it was another non-primary publication (e.g. editorial, technical note); (2) The study reported on (potentially) overlapping study populations.

Data extraction

Data on the study and patients characteristics, the imaging protocol and diagnostic outcomes were extracted by two independent reviewers (D.P.N and R.M.M.) and discrepancies were resolved by consensus. If available, source data (i.e. true positive [TP], false positive [FP], true negative [TN], and false negative [FN]) were extracted or recalculated. If unavailable, the corresponding author of the article was contacted to provide additional data.

Quality assessment

Two authors (D.P.N. and R.M.M.) independently assessed all studies for study quality and discrepancies were resolved by consensus. All included studies were assessed for quality by using the QUality Assessment of studies of Diagnostic Accuracy included in Systematic reviews (QUADAS-2) checklist [19]. The quality of prognostic studies was also assessed with the QUality In Prognostic Studies (QUIPS) checklist [20,21].

Statistical analysis

Diagnostic accuracy data is presented with 95% confidence intervals (95%CI) if presented by the authors, or when we were able to reconstruct a 2 × 2 table. Receiver operating characteristic (ROC) analysis was performed if per-patient data could be extracted using SPSS Statistics (version 20.0; Chicago, IL, USA). The Youden Index (YI) was used to determine the optimal cut-off. P-values were reported as NS (not statistically significant, i.e., $P \geq 0.05$), ≤ 0.05, ≤ 0.01, ≤ 0.001.

Results

The search in Pubmed and Embase retrieved 429 unique studies. After excluding 383 studies on title or abstract we reviewed the full text of 46 studies. Finally, 17 studies were included (10 diagnostic, 5 prognostic and 2 both) for qualitative analysis [22–38] (Fig. 1). Due to heterogeneity in applied b-values, imaging protocols, outcome measurements and reference standards we decided not to perform any quantitative meta-analysis.
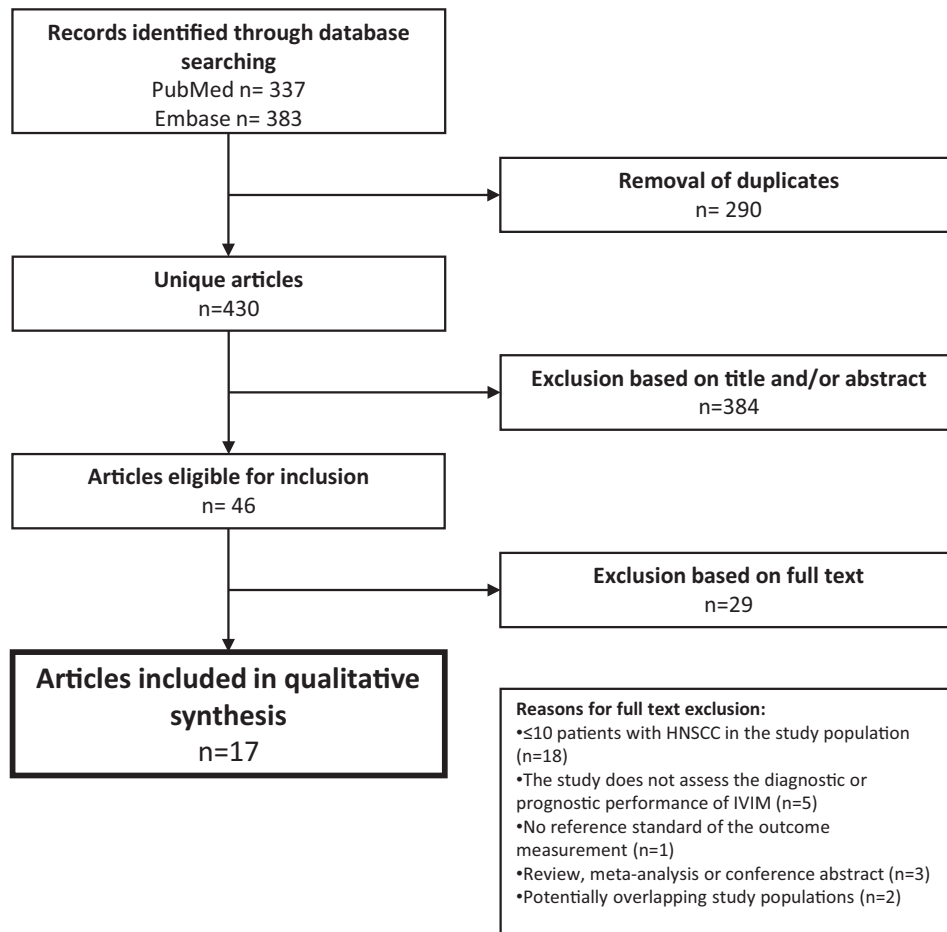
```
┌─────────────────────────────────────┐
│   Records identified through database │
│              searching                │
│          PubMed n= 337                │
│          Embase n= 383                │
└─────────────────────────────────────┘
                 │
                 │────────────────────────►┌──────────────────────────┐
                 │                          │   Removal of duplicates   │
                 │                          │          n= 290           │
                 ▼                          └──────────────────────────┘
┌─────────────────────────────────────┐
│          Unique articles             │
│             n=430                    │
└─────────────────────────────────────┘
                 │
                 │────────────────────────►┌──────────────────────────────────┐
                 │                          │ Exclusion based on title and/or    │
                 │                          │            abstract                │
                 │                          │             n=384                  │
                 ▼                          └──────────────────────────────────┘
┌─────────────────────────────────────┐
│     Articles eligible for inclusion  │
│               n= 46                  │
└─────────────────────────────────────┘
                 │
                 │────────────────────────►┌──────────────────────────┐
                 │                          │ Exclusion based on full text│
                 │                          │          n=29             │
                 ▼                          └──────────────────────────┘
┌─────────────────────────────────────┐
│   Articles included in qualitative   │
│             synthesis                │
│               n=17                   │
└─────────────────────────────────────┘
```

Reasons for full text exclusion:
• ≤10 patients with HNSCC in the study population (n=18)
• The study does not assess the diagnostic or prognostic performance of IVIM (n=5)
• No reference standard of the outcome measurement (n=1)
• Review, meta-analysis or conference abstract (n=3)
• Potentially overlapping study populations (n=2)

Fig. 1. Flow chart of study inclusion.

Study characteristics

In total the selected studies included 882 patients [22–38]. The most common included head and neck malignancies were nasopharyngeal carcinoma (n = 417) and squamous cell carcinoma at other sites (n = 220). Patient characteristics are mentioned in Table 1. In 414 patients with NPC, mean values of D $(0.732 \cdot 10^{-3}$ mm^2/s), D∗ $(84.34 \cdot 10^{-3}$ mm^2/s) and f (18.5%) could be extracted from the articles.

In SCC patients mean data were available for D (n = 160, mean = $0.863 \cdot 10^{-3}$ mm^2/s), D∗ (n = 82, mean = $32.51 \cdot 10^{-3}$ mm^2/s) and f (n = 130, mean = 22.3%). These differences where significant (P < 0.001 for D and D∗ and P = 0.006 for f).

Imaging was performed at either 1.5 T (n = 9 studies) [22,29–34,36,37] or 3 T (n = 8 studies) [23,25–28,36,38,24]. Diffusion-weighted imaging was commonly performed with single-shot spin-echo echo planar imaging (SS-SE-EPI) (n = 13) [23–30,32–38]. In four of these studies the use of fat saturation is mentioned with CHEmical Shift Selective (CHESS) [31], chemical shift-based fat suppression [23] or Spectral Presaturation with Inversion Recovery (SPIR) [27,28]. One study used short tau inversion recovery EPI (STIR-EPI) [22] and another study used Half-fourier acquisition single-shot turbo spin-echo (HASTE) [31]. Diffusion-weighted imaging was acquired with a median of 10.5 b-values (range, 3–17). The used b-values ranges were 0–800 s/mm^2 (n = 7) [23,25,26,30,32,34,38], 0–850 s/mm^2 (n = 1) [29] or 0–1000 s/mm^2 (n = 9) [22,27,28,31,33,35,36,24,37]. Inclusion of low b-values (i.e. b < 200 s/mm^2 [17]) is important for bi-exponential fitting of the signal intensity over b. One study did only include b = 0 s/mm^2 as b-value below b = 200 s/mm^2 [33]. Another study did include 3 b-values below b = 200 s/mm^2 [22]. The other studies included a median of 7 b-values below b = 200 s/mm^2 (range, 4–11) [23,25–32,34–36,38]. The Levenberg-Marquardt algorithm was most commonly used for signal fitting (n = 11) [25–28,31,34–36,38]. In 12 studies it was mentioned that a two-step-fit was used with an estimation of D in the first step with D∗ and f being determined in the second step [23,24,30–34,37,38,26,35,36]. Median scan time was 4:51 min (range 01:30–12:00 min). An overview of imaging characteristics is mentioned in Table 2.

Quality assessment

Results of QUADAS-2 and QUIPS are mentioned in Tables 3 and 4.

The QUADAS-2 yielded the following findings. There was a high risk of a biased patient selection in 6 studies [22,24,26,27,29,32]. In 2 studies this was due to a case-control design [22,27]. In 6 studies there were inappropriate exclusions [22,24,26,27,29,32]. For example by classifying lymph node as benign or malignant based only on imaging criteria, which may result in missing small lymph node metastases [26]; by only including patients with lymph node metastases when the primary tumor is assessed separately, while patients with an N0 neck could have been included as well [24,29]; or by only including patients of whom the tumor was excised, while also including biopsy-proven malignancies would have led to a larger patient population [32]. In none of the 10 studies where a threshold was used this was pre-specified, which may result in an overestimation of diagnostic value [23,24,28,29,31,34–

Table 1
Patient characteristics.

Author, year	Study design	Included patients	Mean age (SD, range)	% Male	Tumor type	AJCC stage	T stage	N stage	M stage	Treatment	Reference standard
Dikaios [22], 2014	...	40	58 (8, 43–79)	...	HNSCC	...	...	Unilateral	...	...	HP
Ding [23], 2015	PS	31	57[a] (44–78)	94	SCC: Tonsil (n = 16), Base of tongue (n = 15)	III (n = 4), IV (n = 27)	...	N0 (n = 1), N1 (n = 3), N2 (n = 17)	...	CRT	RC
Hauser [25], 2013	RS	22	55 (9, 34–69)	73	SCC: Oropharynx (n = 15), Hypopharynx (n = 4), Larynx (n = 2), NPC (n = 1)	IV	...	N0 (n = 1), N1 (n = 1), N2 (n = 18), N3 (n = 2)	...	CRT	RC + CE
Hauser [26], 2014	RS	14	55 (8, 43–69)	60	SCC: Oropharynx (n = 9), Hypopharynx (n = 3), Larynx (n = 3)	...	...	N+	...	CRT	RC + CE
Guo [24], 2016	RS	28	55 (18, 44–72)	100	Hypopharynx	III and IV	...	N2 and N3	M0	IC	RC + HP
Lai [27], 2013	PS	83	52 (12–90)	72	NPC, posttreatment fibrosis	...	T1 (n = 17), T2 (n = 7), T3 (n = 22), T4 (n = 7)	...	...	CRT	HP
Lai [28], 2014	PS	80	51 (14, 12–90)	73	NPC	I (n = 10), II (n = 11), III (n = 44), IV (n = 15)	T1 (n = 33), T2 (n = 7), T3 (n = 31), T4 (n = 9)	N0 (n = 15), N1 (n = 17), N3 (n = 5)	M0 (n = 76), M1 (n = 4)	...	HP
Lu [29], 2013	RS	16	55 (38–64)	94	SCC: Oropharynx (n = 11), Oral cavity (n = 4), NPC (n = 1)	III (n = 1), IV (n = 15)	...	N+	...	Surgery (n = 2), CRT (n = 14)	HP
Marzi [30], 2013	RS	37	57 (30–76)	84	SCC: Oropharynx (n = 15), NPC (n = 11), Hypopharynx/Larynx (n = 8), Oral Cavity (n = 2), Maxillary sinus (n = 1),	...	T1 (n = 3), T2 (n = 13), T3 (n = 8), T4 (n = 13)	N0 (n = 3), N1 (n = 3), N2 (n = 28), N3 (n = 3)	...	...	HP
Sakamoto [31], 2014	PS	33	59 (14–89)	48	Basaloid SCC (n = 1), Carcinoma ex pleomorphic adenoma (n = 1), adenoid cystic carcinoma (n = 1), SCC (n = 18), Verrucous carcinoma (n = 2)	...	...	...	...	...	HP
Sasaki [32], 2014	RS	94	62 (15, 3–91)	60	Various benign and malignant head and neck tumors	...	...	...	...	...	HP
Sumi [33], 2012	RS	113	60 (21–91)	55	Various benign and malignant head and neck tumors	...	...	...	...	...	HP
Sumi [34], 2012	PS	31	61 (21–82)	52	SG: Benign (n = 20), Malignant (n = 11)	...	...	...	...	Surgery	HP
Xiao [35], 2015	PS	48	42[a] (13–65)	69	NPC	III (n = 19), IV (n = 29)	T1 (n = 5), T2 (n = 7), T3 (n = 14), T4 (n = 22)	N0 (n = 3), N1 (n = 9), N2 (n = 25), N3 (n = 11)	M0 (n = 45), M1 (n = 3)	NAC	RC
Xiao-Ping [36], 2015	PS	50	49 (11)	78	NPC	III (n = 12), IV (n = 38)	T1 (n = 1), T2 (n = 10), T3 (n = 19), T4 (n = 20)	N0 (n = 4), N1 (n = 9), N2 (n = 17), N3 (n = 20)	M0 (n = 31), M1 (n = 19)	NAC + (C) RT	RC
Yu [37], 2016	RS	102	...	...	NPC (n = 80), Lymphoma (n = 22)	...	T1 (n = 8), T2 (n = 29), T3 (n = 19), T4 (n = 26)	...	...	...	HP
Zhang [38], 2014	PS	60	51 (16–69)	74	NPC	...	...	...	...	...	HP

Abbreviations: AJCC = American Joint Committee on Cancer; CE = clinical evaluation; CRT = chemoradiotherapy; HNSCC = head and neck squamous cell carcinoma; HP = histopathology; IC = induction chemotherapy; NAC = neoadjuvant chemotherapy; NPC = nasopharyngeal carcinoma; PS = prospective; RC = RECIST; RS = retrospective; RT = radiotherapy; SCC = squamous cell carcinoma; SG = salivary gland tumor.

[a] Median age.

Table 2
Imaging characteristics.

Author, year	Field strength (T)	Diffusion sequence	TR/TE (ms)	FOV (mm)	Matrix	Section thickness (mm)	b-values	IVIM fit	IVIM parameters	Scan time
Dikaios [22], 2014	1.5	STIR-EPI	8700/88	200	128 × 128	4	0, 50, 100, 300, 600, 1000	MP, NR	D, D*, f	06:10
Ding [23], 2015	3	SS-SE-EPI	3600/100	256	128 × 128	3.5	0, 20, 40, 60,80, 100, 120, 150, 200, 400, 600, 800	BeF, SLF, NLLS	D, D*, f	…
Guo [24], 2016	3	SS-SE-EPI	2500/79	230	256 × 256	5	0, 10, 20, 30, 50, 70, 100, 150, 200, 400, 800, 1000	BeF using LMa	D, D*, f	05:08
Hauser [25], 2013	3	SS-SE-EPI	1300/50	240	80 × 80	3	0, 50, 100, 150, 200, 250, 700, 800	BeF using LMa	D, f	04:51
Hauser [26], 2014	3	SS-SE-EPI	1300/50	240	80 × 80	3	0, 50, 100, 150, 200, 250, 700, 800	BeF using LMa	D, f	04:51
Lai [27], 2013	3	SS-SE-EPI	7996/43	230	256 × 256	3	0, 10, 20, 30, 40, 60, 100, 120, 160, 200, 300, 500, 1000	BeF using LMa	D, D*, f	12:00
Lai [28], 2014	3	SS-SE-EPI	7996/43	230	256 × 256	3	0, 10, 20, 30, 40, 60, 100, 120, 160, 200, 300, 500, 1000	BeF using LMa	D, D*, f	12:00
Lu [29], 2013	1.5	SS-SE-EPI	4000/90	200–220	128 × 128	6–8	0, 13, 17, 23, 30, 40, 53, 70, 92, 122, 161, 212, 280, 369, 488, 644, 850	BeF using NLLS	D, D*, f	04:00
Marzi [30], 2013	1.5	SS-SE-EPI	4500/77	260–280	128 × 128	4	0, 25, 50, 75, 100, 150, 300, 500, 800	BeF	D, D*, f, f_{asym}	06:13
Sakamoto [31], 2014	1.5	HASTE-DWI	3000/101	230 × 173	192 × 144	4–5	0, 20, 40, 60, 80, 100, 150, 200, 500, 1000	BeF using LMa	D, D*, f	01:30
Sasaki [32], 2014	1.5	SS-SE-EPI	1625/81	200	112 × 90	4	0, 10, 20, 30, 50, 80, 100, 200, 300, 400, 800	Geo, LS	D, D*, f, P	01:53
Sumi [33], 2012	1.5	SS-SE-EPI	4283/87	200	112 × 90	4	0, 500, 1000	SIF	D, PP	02:08
Sumi [34], 2012	1.5	SS-SE-EPI	1625/81	200	112 × 90	4	0, 10, 20, 30, 50, 80, 100, 200, 300, 400, 800	BeF using LMa	D, D*, f	01:53
Xiao [35], 2015	3	SS-SE-EPI	4495/69	230	256 × 256	5	0, 10, 20, 30, 40, 50, 100, 150, 200, 350, 500, 650, 800, 1000	BeF using LMa	D, D*, f	06:00
Xiao-Ping [36], 2015	1.5	SS-SE-EPI	4225/106	220	128 × 130	5	0, 50, 80, 100, 150, 200, 400, 600, 800, 1000	BeF using LMa	D, D*, f	02:53
Yu [37], 2016	1.5	SS-SE-EPI	4225/106	220	128 × 130	5	0, 50, 80, 100, 150, 200, 400, 600, 800, 1000	BeF using LMa	D, D*, f	…
Zhang [38], 2014	3	SS-SE-EPI	3000/58	240	128 × 128	4	0, 10, 20, 30, 50, 80, 100, 150, 200, 300, 400, 600, 800	BeF using LMa	D, D*, f	03:45

Abbreviations: BeF = bi-exponential fit; D* = pseudodiffusion coefficient; D = diffusion coefficient; f = perfusion factor; f_{asym} = perfusion factor estimated using an asymptotic method; Geo = geometric method; HASTE DWI = half-Fourier acquisition single-shot turbo spin-echo diffusion-weighted imaging; LMa = Levenberg Marquard algorithm; LS = least squares method; MD = median model; MP = maximum probability model; NLLS = non-linear least-squares; NR = nonlinear regression model; P = perfusion parameter that is heavily weighted towards extravascular space; PS = prospective; PP = perfusion-related parameter; RS = retrospective; SS-SE-EPI = single-shot spin-echo echo planar imaging; SIF = simplified IVIM formula; STIR-EPI = short tau inversion recovery echo planar imaging; SLF = simplified linear fit.

38]. In 2 studies the observers were not blinded to the reference standard [29,30] and in 11 studies this was unclear [22,24–28,32–34,37,38]. In 5 studies imaging was used as reference standard instead of histopathology [22,23,25,35,36]. The interval between index test and reference standard was not mentioned in 10 studies [22,27–34,36].

For the prognostic studies the assessment with QUIPS yielded the following results. In 4 studies reasons for loss to follow-up are mentioned without describing participants who were lost to follow-up [23,25,26,36]. Xiao et al. excluded patients with insufficient follow-up images [35]. None of the included studies attempted to correct for possible confounders [23–26,29,35,36]. Lu et al. used both parametric and non-parametric statistical tests on the same data which raised concerns on the adequacy of the

statistical analysis [29]. In 3 studies we were unable to reconstruct a two-by-two table based on the given sensitivity and specificity [24,35,36].

Diagnostic study results

Study results are mentioned in Table 5 and Appendix B. In 9 studies the primary tumor was assessed separately [24,27,28,30–32,34,37,38], in 2 studies lymph nodes were assessed separately [22,28] and in 4 studies a combination of primary tumor and lymph nodes was assessed [28,29,32,33].

Sasaki et al. [32] provided an appendix with per-patient IVIM data. The authors used both the aforementioned IVIM formula

Table 3
Results of QUADAS-2 for bias assessment of all included studies: '✔' indicates a low risk of bias; '?' an unclear risk and 'x' indicates a high risk of bias.

	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Dikaios [22], 2014	x	?	?	?	?	✔	✔
Ding [23], 2015	?	x	?	✔	✔	✔	✔
Guo [24], 2016	x	x	✔	✔	✔	✔	✔
Hauser [25], 2013	✔	?	?	x	✔	✔	✔
Hauser [26], 2014	x	?	✔	x	✔	✔	✔
Lai [27], 2013	x	?	✔	✔	x	✔	✔
Lai [28], 2014	✔	x	✔	?	✔	✔	✔
Lu [29], 2013	x	x	✔	✔	✔	✔	✔
Marzi [30], 2013	?	x	✔	x	✔	✔	✔
Sakamoto [31], 2014	?	x	✔	?	✔	✔	✔
Sasaki [32], 2014	x	?	✔	✔	✔	✔	✔
Sumi [33], 2012	?	?	✔	✔	✔	✔	✔
Sumi [34], 2012	✔	x	✔	✔	✔	✔	✔
Xiao [35], 2015	✔	x	?	✔	✔	✔	✔
Xiao-Ping [36], 2015	✔	x	?	✔	✔	✔	✔
Yu [37], 2016	?	x	✔	✔	✔	✔	✔
Zhang [38], 2014	✔	x	✔	✔	✔	✔	✔

Table 4
QUIPS results for bias assessment of prognostic studies: '✔' indicates a low risk of bias; '?' a moderate risk and 'x' indicates a high risk of bias.

	1. Study participation	2. Study attrition	3. Prognostic factor measurement	4. Outcome measurement	5. Study confounding	6. Statistical analysis and reporting
Ding [23], 2015	✔	?	?	?	x	✔
Guo [24], 2016	x	✔	?	?	x	x
Hauser [25], 2013	✔	x	?	x	x	?
Hauser [26], 2014	x	?	x	?	x	?
Lu [29], 2013	✔	x	?	x	x	x
Xiao [35], 2015	✔	?	?	?	x	x
Xiao-Ping [36], 2015	✔	?	?	?	x	x

and an IVIM analysis based on a geometric method using only 3 b-values:

$$GeoD = \frac{(\ln S_{200} - \ln S_{800})}{600}$$

$$Geof = \frac{1 - S_{inter}}{S_0}$$

$$GeoP = \frac{(\ln S_0 - \ln S_{inter})}{200}$$

S_{inter} is the interception of the logarithmic regression line obtained using b-values of 200 and 800 s/mm^2 with the y-axis [32]. Because the GeoP is fundamentally different than D∗ the authors used another symbol; GeoP is more weighted to the extravascular space compared to D∗ which is more weighted towards the vascular space. Based on the data provided in the appendix we used ROC analysis to determine the diagnostic accuracy of IVIM (with both methods) to discriminate SCC from lymphoma as these were the most prevalent malignancies in this study [32]. Further we compared IVIM values between primary SCC and SCC lymph node metastases from Sasaki et al. [32] to validate the results of Lu et al. [29] of which the results will be discussed below.

The abilities of IVIM to differentiate between various benign and malignant head and neck lesions were determined by five studies [31–34,37]. Sakamoto et al. [31] found D to be the single most valuable parameter with a sensitivity of 87% (95%CI, 66–97%) and specificity of 80% (95%CI, 44–97%). When combining D and D∗ sensitivity and specificity increased to 91% (95%CI, 72–

99%) and 90% (95%CI, 56–100%) respectively. Included benign lesions were mainly pleomorphic adenomas (4/10) and vascular malformations (4/10) and malignant lesions were predominantly SCCs (18/23). Yu et al. focused on differentiating between NPC and lymphoma [37]. All IVIM values (D, D∗ and f) were significantly lower in lymphoma. The highest diagnostic accuracy was achieved when D∗ and f were combined with sensitivity and specificity being 85% (95%CI, 76–92%) and 100% (95%CI, 83–100%), respectively. These results are in line with Sumi et al. [33] who reported that IVIM values (D and the perfusion-related parameter [PP]) differed significantly ($P \le 0.001$) between six types of head and neck lesions (lymphoma, SCC, malignant salivary gland (SG) tumor, Warthin tumor, pleomorphic adenoma and schwannoma). Lymphomas had significantly lower D and PP values than SCC ($P \le 0.001$). Malignant SG tumor D was significantly lower than D of pleomorphic adenoma and significantly higher than D of Warthin tumors ($P \le 0.01$). These results were verified in two other studies [32,34]. In all three studies malignant SG tumors had intermediate D and D∗ values compared to Warthin tumors (lower values) and pleomorphic adenoma (higher values). The combined use of D and D∗ could separate malignant from benign SG tumors with a sensitivity of 100% (95%CI, 54–100%) and a specificity of 94–100% (95%CI, 71–100%) [32,34].

Marzi et al. [30] compared pretreatment IVIM values of different SCC locations: nasopharynx (n = 11), oropharynx (n = 15) and hypopharynx/larynx (n = 8). Both D ($P \le 0.01$) and f ($P \le 0.05$) were significantly different between groups. D was highest in hypopharyngeal/laryngeal SCC (D_{median} = 1.07 · 10^{-3} mm^2/s) and lowest in nasopharyngeal SCC (D_{median} = 0.83 · 10^{-3} mm^2/s). For f, oropharyngeal SCC had the highest values (f_{median} = 22.5% vs < 18.6%)

Table 5
Diagnostic accuracy.

	n	Parameter	Outcome measurement	Cut-off	Sensitivity (95%CI)	Specificity (95%CI)	AUC (95%CI)
Primary tumor							
Lai [27], 2013	83	D	NPC PT vs posttreatment fibrosis	$1.06 \cdot 10^{-3}$ mm²/s	100 (93–100)	100 (88–100)	1.00 (1.00–1.00)
Lai [27], 2013	83	D*	NPC PT vs posttreatment fibrosis	$85.3 \cdot 10^{-3}$ mm²/s	100 (93–100)	91 (76–98)	0.99 (0.97–1.00)
Lai [27], 2013	83	f	NPC PT vs posttreatment fibrosis	0.13	66 (52–79)	100 (82–100)	0.89 (0.82–0.96)
Lai [28], 2014	80	D	NPC T-stage: low vs high	$0.76 \cdot 10^{-3}$ mm²/s	83 (68–93)	53 (36–69)	0.65
Lai [28], 2014	80	f	NPC T-stage: low vs high	0.13	81 (65–91)	93 (80–98)	0.90 (0.82–0.96)
Lai [28], 2014	80	D*	NPC T-stage: low vs high	$101.0 \cdot 10^{-3}$ mm²/s	73 (57–86)	83 (67–93)	0.83
Sakamoto [31], 2014	33	D + D*	Malignant vs benign	…	91 (72–99)	90 (56–100)	0.96 (0.90–1.00)
Sakamoto [31], 2014	33	D	Malignant vs benign	$0.98 \cdot 10^{-3}$ mm/s²	87 (66–97)	80 (44–97)	0.91 (0.81–1.00)
Sakamoto [31], 2014	33	f	Malignant vs benign	0.19	74 (52–90)	50 (19–81)	0.52 (0.29–0.74)
Sakamoto [31], 2014	33	D*	Malignant vs benign	$8.42 \cdot 10^{-3}$ mm²/s	61 (39–80)	90 (56–100)	0.75 (0.58–0.97)
Sasaki [32], 2014	35	Fit D	SCC vs lymphoma	$0.84 \cdot 10^{-3}$ mm²/s	74 (56–87)	100 (74–100)	0.87 (0.77–0.97)
Sasaki [32], 2014	35	Geo D	SCC vs lymphoma	$0.87 \cdot 10^{-3}$ mm²/s	74 (56–87)	100 (74–100)	0.87 (0.76–0.97)
Sasaki [32], 2014	35	Fit f	SCC vs lymphoma	0.11	50 (32–68)	83 (52–98)	0.61 (0.45–0.78)
Sasaki [32], 2014	35	Geo P	SCC vs lymphoma	$0.57 \cdot 10^{-3}$ mm²/s	50 (32–68)	83 (52–98)	0.61 (0.44–0.78)
Sasaki [32], 2014	35	Geo f	SCC vs lymphoma	0.11	50 (32–68)	83 (52–98)	0.61 (0.44–0.78)
Sasaki [32], 2014	23	Geo D + Geo P	SG malignant vs benign	…	100 (54–100)	94 (71–100)	…
Sasaki [32], 2014	23	Geo D + Geo f	SG malignant vs benign	…	100 (54–100)	94 (71–100)	…
Sasaki [32], 2014	23	Fit D + Fit D*	SG malignant vs benign	…	100 (54–100)	94 (71–100)	…
Sasaki [32], 2014	23	Fit D + Fit f	SG malignant vs benign	…	50 (12–88)	100 (80–100)	…
Sasaki [32], 2014	23	Geo D₄b + Geo P₄b	SG malignant vs benign	…	83 (36–100)	88 (64–99)	…
Sasaki [32], 2014	23	Geo D₄b + Geo f₄b	SG malignant vs benign	…	83 (36–100)	88 (64–99)	…
Sumi [34], 2012	31	D*	SG malignant vs benign	$10 < D^* < 23 \cdot 10^{-3}$ mm²/s	73 (39–94)	65 (41–85)	…
Sumi [34], 2012	31	D	SG malignant vs benign	$0.8 < D < 1.1 \cdot 10^{-3}$ mm²/s	64 (31–89)	100 (83–100)	…
Sumi [34], 2012	31	D + D*	SG malignant vs benign	…	100 (72–100)	100 (83–100)	…
Sumi [34], 2012	31	D + f	SG malignant vs benign	…	82 (48–98)	100 (83–100)	…
Sumi [34], 2012	31	D* + f	SG malignant vs benign	…	82 (48–98)	65 (41–85)	…
Yu [37], 2016	102	D	NPC vs lymphoma	$0.66 \cdot 10^{-3}$ mm²/s	55 (44–66)	100 (83–100)	0.80 (0.71–0.87)
Yu [37], 2016	102	D*	NPC vs lymphoma	$7.89 \cdot 10^{-3}$ mm²/s	83 (73–90)	85 (62–97)	0.90
Yu [37], 2016	102	f	NPC vs lymphoma	0.29	41 (31–53)	95 (75–100)	0.66 (0.54–0.74)
Yu [37], 2016	102	fD*	NPC vs lymphoma	$1.99 \cdot 10^{-3}$ mm²/s	85 (76–92)	100 (83–100)	0.96 (0.90–0.99)
Zhang [38], 2014	60	D	PT vs adenoid hypertrophy	$0.75 \cdot 10^{-3}$ mm²/s	83	65	0.85
Lymph node							
Dikaios [22], 2014	40	mp-DWI	SCC NR benign vs metastatic	…	80	94	0.95 (0.80–1.00)
Dikaios [22], 2014	40	mp-DWI	SCC NR benign vs metastatic	…	80	88	0.92 (0.85–1.00)
Lai [28], 2014	80	D	NPC N stage: low vs high	$0.76 \cdot 10^{-3}$ mm/s²	88 (75–95)	66 (48–82)	0.86 (0.76–0.93)
Lai [28], 2014	80	f	NPC N stage: low vs high	0.15	85 (72–94)	61 (42–77)	0.69
Lai [28], 2014	80	D*	NPC N stage: low vs high	$103.9 \cdot 10^{-3}$ mm²/s	71 (56–83)	55 (36–72)	0.63
Primary tumor + lymph node							
Lai [28], 2014	80	D	NPC low vs high AJCC stage	$0.78 \cdot 10^{-3}$ mm/s²	93 (84–98)	76 (53–92)	0.91 (0.83–0.97)
Lai [28], 2014	80	f	NPC low vs high AJCC stage	0.15	88 (77–95)	86 (64–97)	0.87
Lai [28], 2014	80	D*	NPC low vs high AJCC stage	$100.4 \cdot 10^{-3}$ mm²/s	48 (35–62)	90 (70–99)	0.72
Lu [29], 2013	16	f + D	SCC PT vs LN	…	63 (35–85)	81 (54–96)	0.76
Lu [29], 2013	16	f	SCC PT vs LN	0.22	63 (35–85)	75 (48–93)	0.71
Lu [29], 2013	16	D*	SCC PT vs LN	$43.2 \cdot 10^{-3}$ mm²/s	63 (35–85)	63 (35–85)	0.53
Lu [29], 2013	16	D	SCC PT vs LN	$0.80 \cdot 10^{-3}$ mm²/s	56 (30–80)	94 (70–100)	0.74
Sasaki [32], 2014	34	Fit f	SCC PT vs LN	0.06	100 (85–100)	27 (6–61)	0.62 (0.41–0.84)
Sasaki [32], 2014	34	Geo f	SCC PT vs LN	0.05	96 (78–100)	27 (6–61)	0.53 (0.31–0.75)
Sasaki [32], 2014	34	Geo P	SCC PT vs LN	$0.26 \cdot 10^{-3}$ mm²/s	96 (78–100)	27 (6–61)	0.53 (0.31–0.75)

Abbreviations: AJCC = American Joint Committee on Cancer; AUC = area under the curve; D* = pseudodiffusion coefficient; D = diffusion coefficient; D_{4b} = diffusion coefficient acquired with four b-values; f = perfusion factor; f_{4b} = perfusion factor acquired with four b-values; Fit = least-squares method; Geo = geometric method; LN = lymph node; mp-DWI = multiparametric analysis of diffusion-weighted imaging parameters; NPC = nasopharyngeal carcinoma; NR = nonlinear regression; P = perfusion parameter that is heavily weighted towards extravascular space; P_{4b} = perfusion parameter that is heavily weighted towards extravascular space acquired with four b-values; PT = primary tumor; SCC = squamous cell carcinoma; SG = salivary gland tumors; 95%CI = 95% confidence interval.

and also the largest range in values (f_{range} = 12.6–32.7%). Nasopharyngeal SCCs were the most homogeneous group (f_{range} = 7.3–16.7%).

Two studies assessed the difference of IVIM parameters between primary SCC and lymph nodes. Lu et al. [29] showed that lymph nodes have significantly higher D values ($P \leq 0.001$) and lower f values ($P \leq 0.001$). These results could not be confirmed by Sasaki et al. [32] who did not find any significant differences between primary SCC and SCC lymph nodes. It should be noted that in the first study included patients all had a primary tumor and a lymph node metastasis; whereas in the second study it is not specified whether primary tumors and lymph node metastasis of the same patients were assessed.

The differentiation between malignant tumor and post-chemoradiation fibrosis was assessed by Lai et al. [27] in a case-control study including pre-treatment NPC and biopsy-confirmed post-chemoradiation fibrosis. With D it was possible to separate both with a sensitivity (95%CI, 93–100%) and specificity (95%CI, 88–100%) of 100%.

In another study by Lai et al. [28] NPC at different disease stages were compared. For T stage, N stage and disease stage according to the American Joint Committee on Cancer (AJCC), all IVIM parameters (i.e., D, D∗ and f) were significantly lower in the high stage group than in the low stage group. For N staging and AJCC staging, D had the highest diagnostic accuracy (AUC = 0.86 and 0.91, respectively), while f had the highest AUC in T staging

(AUC = 0.90). In multivariate analysis there was a non-significant trend towards a higher AUC for the combined use of all IVIM parameters.

Dikaios et al. [22] used a (conventional) non-linear (NR) and a maximum probability (MP) model to estimate IVIM-parameters for differentiating between benign and malignant lymph nodes. No significant differences between models were found. Both f and D_* were significantly different ($P \leq 0.01$) between benign and malignant nodes while D was not.

Prognostic study results

Prognostic study results are mentioned in Table 6 and Appendix C. In 5 studies the primary tumor was assessed [23–25,29,35,36] separately, in 3 the lymph nodes were assessed separately [23,26,35] and in one both were assessed simultaneously [23].

The value of IVIM in predicting response to neo-adjuvant chemotherapy (NAC) in NPC was assessed by Xiao et al. [35] and Xiao-Ping et al. [36]. In both studies pre-treatment D could significantly predict the response to NAC ($P \leq 0.01$) with a sensitivity and specificity of 64–65% and 72–81%, respectively. In both studies f was a weaker predictor than D, with only Δf being a significant predictor in the study of Xiao et al. (primary tumor: $P \leq 0.05$; lymph node: $P \leq 0.01$) [35].

In the study of Xiao-Ping et al. sensitivity increased to 94% and specificity was 77% when the difference in D before and after NAC was used to assess the effect of NAC [36]. It should be noted that pre-NAC D had a relatively high diagnostic accuracy in predicting the presence of residual disease after CRT with a sensitivity of 82% and a specificity of 83%. Ding et al. performed IVIM before and during CRT in HPV-positive primary HNSCC and found pre-treatment D to differ significantly between responders and non-responders [23]. Guo et al. found similar results in predicting treatment response in NAC for hypopharyngeal SCC with D being the

strongest predictor with a sensitivity of 75% and specificity being 89% [24].

These findings are in contrast to Hauser et al. who performed one study on primary tumors [25] and one on lymph nodes [26]. The patient population consisted of HNSCC patients receiving CRT. In both studies pretreatment f discriminated best between responders and non-responders ($P \leq 0.01$), whereas D was not statically significant.

To conclude it should be noted that in only 1 of the prognostic studies D_* differed statistically significantly between responders and non-responders [24].

Discussion

In studies on patients with SG tumors, all IVIM values of malignant SG tumors were between those of Warthin tumors and pleomorphic adenomas [32–34]. This demonstrates that diagnostic accuracy of IVIM in distinguishing between malignant and benign head and neck lesions depends strongly on the included tumor types. With IVIM it was possible to reliably differentiate between SCCs, lymphomas, malignant SG tumors, Warthin tumors and pleomorphic adenomas [31,33,34,37]. It should be noted that combining IVIM parameters often yielded a higher diagnostic accuracy than using a single IVIM parameter [22,28,29,31,32,34,37]. Future research should focus on finding the optimal combination of functional imaging parameters, for example by combining IVIM with dynamic contrast-enhanced MRI [39]. Preferably only lesions should be included which are currently challenging to separate.

Detection of residual disease in irradiated tissue remains challenging [40]. A recent randomized trial did reveal that ^{18}F-FDG-PET/CT can reduce the need for investigations under anesthesia after RT for laryngeal carcinoma without compromising treatment quality [41]. Especially early after (chemo) radiotherapy inflammation may result in residual ^{18}F-FDG uptake. Combined with a rela-

Table 6
Prognostic accuracy.

	n	Follow-up	Tumor type	Parameter	Outcome measurement	Cut-off	Sensitivity (95%CI)	Specificity (95%CI)	AUC (95%CI)
Primary tumor									
Guo [24], 2016	28	3 weeks	Hypopharynx SCC	Pre-treatment D	Response to NAC	$0.85 \cdot 10^{-3}$ mm²/s	75	89	0.81
Hauser [25], 2013	22	≥7.5 months	SCC/NPC	Pre-treatment D	Locoregional treatment outcome	...	83 (36–100)	81 (54–96)	0.79 (0.58–1.00)
Hauser [25], 2013	22	≥7.5 months	SCC/NPC	Pre-treatment f	Locoregional treatment outcome	...	67 (22–96)	94 (70–100)	0.82 (0.62–1.00)
Xiao [35], 2015	48	...	NPC	Pre-treatment D	Response to NAC	$0.91 \cdot 10^{-3}$ mm²/s	64	81	0.71
Xiao-Ping [36], 2015	50	...	NPC	$\Delta D_{pre-post\ NAC}$	Effect of NAC[b]	26.3%	94	77	0.86 (0.72–1.00)
Xiao-Ping [36], 2015	50	...	NPC	Pre-treatment D	Residue after CRT[a]	$0.73 \cdot 10^{-3}$ mm²/s	82	83	0.84 (0.70–0.98)
Xiao-Ping [36], 2015	50	...	NPC	Pre-treatment D	Effect of NAC[b]	$0.95 \cdot 10^{-3}$ mm²/s	65	72	0.77 (0.61–0.92)
Xiao-Ping [36], 2015	50	...	NPC	$\Delta D_{pre-post\ NAC}$	Residue after CRT[a]	25.0%	79	83	0.76 (0.59–0.92)
Lymph node									
Hauser [26], 2014	14	≥13.5 months	SCC	Pre-treatment D	Locoregional treatment outcome	...	100 (29–100)	55 (23–83)	0.71 (0.40–1.00)
Hauser [26], 2014	14	≥13.5 months	SCC	Pre-treatment f	Locoregional treatment outcome	...	100 (29–100)	100 (72–100)	1.00 (1.00–1.00)
Xiao [35], 2015	48	...	NPC	Pre-treatment D	Response to NAC	$0.95 \cdot 10^{-3}$ mm²/s	55	95	0.77

Abbreviations: AUC = area under the curve; CRT = chemoradiotherapy; D = diffusion coefficient; f = perfusion factor; NAC = neoadjuvant chemotherapy; NPC = nasopharyngeal carcinoma; SCC = squamous cell carcinoma; 95%CI = 95% confidence interval.
[a] The authors did not specify the interval between imaging and the assessment of residual disease after treatment.
[b] The authors considered treatment effective if patients exhibited a complete response (CR) or partial response (PR) based on the RECIST criteria.

tively low prevalence of residual disease this results in a relatively poor positive predictive value (PPV) of PET-CT, which leaves room for improvement [40]. Performing additional IVIM imaging may enhance the PPV of ^{18}F-FDG-PET/CT. Lai et al. [27] showed that pretreatment NPC could be separated from posttreatment fibrosis with an accuracy of 100%. In practice residual tumor foci may be hidden in fibrotic tissue and therefore may be more challenging to detect. Further research should assess if the resolution of IVIM imaging is appropriate to detect residual disease in the proximity to fibrosis with a high diagnostic accuracy. In order to identify the most optimal timing of IVIM imaging after irradiation more research is necessary. Preferably by performing IVIM multiple times during and after radiation therapy. Besides individual IVIM values, it may also be valuable to look at Δ values between different time points. In the studies included in this systematic review follow-up was relatively short, with a maximum of 3 months [23,24], in the studies where intra- or posttreatment imaging was performed.

As shown by Marzi et al. [30] nasopharyngeal tumors have lower D and f values than SCC at other locations. When we pooled the results of the individual studies reporting mean values for D, D∗ and f, we could confirm the results of Marzi et al. This difference in IVIM characteristics is another argument to include only one tumor site in future studies, or at least to perform subgroup analyses when both nasopharyngeal carcinoma and other HNSCC are included.

In only one of the prognostic studies D∗ was significantly different between responders and non-responders suggesting that this parameter has the least potential to predict prognosis [24]. There is less consensus on the prognostically most promising parameter. In 4 studies D had the highest predictive value [23,24,35,36] and in 2 studies f was the strongest predictor of treatment outcome [25,26].

A proposed hypothesis is that high f is associated with a higher regional blood flow [42]. High regional blood flow may be indicative of a high microvessel density which is associated with a higher likelihood of both lymph node and distant metastases [43]. Therefore high f could be an unfavorable prognostic factor for survival.

D appears to be inversely correlated with cell density [44,45]. Highly cellular tumors with rapidly dividing cells are more sensitive to chemotherapy and radiotherapy and therefore associated with a more favorable prognosis [23,24,35,36]. An increase in D during therapy is therefore a sign of decreasing cellularity and a good response to treatment [23,24,35,36].

Lai et al. [28] showed that IVIM values differed between disease stages (AJCC, T stage and N stage). Tumors with a high cell turnover with more cell division may result in a larger tumor with densely packed cells resulting in both a higher disease stage and lower D values. For more perfusion weighted parameters (D∗ and f) it can be hypothesized that larger tumors are more prone to intratumoral necrosis due to lacking vascularization in the central portion of the tumor and therefore lower D∗ and f values. This is may have implications for future prognostic studies, because it warrants the need for correcting for disease stage when assessing the prognostic value of IVIM in HNC.

Imaging protocol

The optimal combination and number of b-values remains one of the keys points which need to be addressed. Lemke et al. [46] concluded that at least 10 b-values should be used for fitting the IVIM signal in clinical settings in a simulation study on abdominal IVIM imaging while assigning more weight to low b-values. Most b-values should be in the low range (b = 0–100 s/mm^2) with only a few b-values of >450 s/mm^2 are necessary [46]. This is in contrast to the findings of Gurney-Champion et al. who found 7 b-values to

be enough for abdominal imaging and only 3 b-values if only the liver is of interest in a study on 16 healthy volunteers [47].

In none of the included studies imaging was performed on both a 1.5 T and 3 T MRI system on the same patients. Therefore we could not determine whether the field strength matters for IVIM in head and neck imaging. We do consider this to be an important issue for further research. If 1.5 T and 3 T IVIM data prove to be comparable this would create more opportunities for multicenter research. In abdominal imaging there is some evidence suggesting that D and f values of the liver are reproducible on both 1.5 T and 3 T, whereas D∗ values are more variable [48]. However at this moment DWI values are not considered to be robust enough to be interchangeable between institutions, regardless of field strength [49].

In the head and neck area Sasaki et al. compared the traditional least-squares method requiring 11 b-values with a geometric approach requiring only 3 b-values for differentiating between tumor types [32]. Even though D values were significantly higher and f values significantly lower with the geometric approach compared to the least-squares method; diagnostic accuracy was comparable for differentiating SCC from lymphoma and characterizing SG tumors.

We recommend to include at least 4 b-values below b = 200 s/mm^2 in order to appropriately fit estimated perfusion-related parameters (i.e. D∗, f, P and PP) until a proposed method which requires fewer b-values is appropriately validated [46]. Two of the included studies did not fulfill this criterion [22,33]. The reported D∗, f and PP values of these studies should therefore be interpreted with caution.

Therefore future research should focus on the optimal combination of b-values as well as on the optimal model for determining IVIM parameters. Especially for fitting D∗ and f with IVIM it is important that imaging is performed with a high signal-to-noise ratio to avoid biased parameters [22]. Furthermore it should be noted that the b-value represents the strength of the diffusion pulse and is dependent of the used gradient pulse sequence, the gradient pulse duration and the gradient strength [50]. It is therefore possible to create identical b-values with different imaging parameters which may result in differences in ADC values [51]. More uniform definitions of b-values may improve comparability between protocols.

D∗ appears to be the least robust IVIM parameter with reported mean D∗ values of SCC ranging from $4 \cdot 10^{-3}$ mm^2/s to $49 \cdot 10^{-3}$ mm^2/s [22,24,29,32] and for NPC ranging from $18 \cdot 10^{-3}$ mm^2/s to $153 \cdot 10^{-3}$ mm^2/s [27,28,37,38].

Limitations

Even though this review provides an extensive overview of the use of IVIM imaging in HNC, there are some limitations. Firstly, the included studies were heterogeneous in applied b-values, imaging protocols, tumor types and outcome measurements. This prevented us from performing any meaningful meta-analysis. The differences in selection and number of b-values in the lower range may compromise the comparability of results. Secondly, the relatively small population size comprised the analysis of confounders in the prognostic studies. The outcome parameters, mainly in prognostic studies, were also heterogeneous. It would be preferable if all prognostic studies reported at least one uniform outcome measurement. Even if this standard may not be a perfect gold standard, for example the RECIST criteria [52]. Thirdly, most studies did not correct for multiple testing which may overestimate the number of significant findings. Given that most studies are relatively small and that all can be regarded as positive studies it is likely that publication bias is present. Small studies with negative results may have been regarded to be not interesting enough for publication.

The heterogeneity of the included studies, especially in terms of comparisons made and outcome measurements, made it impossible to perform statistical testing for the presence of publication bias.

Conclusions

With this systematic review we provide an overview of studies on IVIM in HNC. Studies are very heterogeneous in terms of applied b-values, imaging protocols, outcome measurements and reference standards. With combinations of IVIM-parameters SCC, lymphoma, malignant SG tumors, Warthin tumors and pleiomorphic adenoma can be reliably separated from each other. Low pre-treatment D and f and an increase in D during treatment was associated with a favorable response to treatment. D∗ appears to be the parameter with the lowest prognostic value. Future research should focus on finding the optimal IVIM protocol. When assessing diagnostic and prognostic properties of IVIM, authors should use uniformly accepted study methods and larger patient populations.

Conflict of interest statement

None declared.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.oraloncology.2017.03.016.

References

[1] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin 2016;66:7–30.
[2] Gatta G, Botta L, Sanchez MJ, Anderson LA, Pierannunzio D, Licitra L, et al. Prognoses and improvement for head and neck cancers diagnosed in Europe in early 2000s: the EUROCARE-5 population-based study. Eur J Cancer 2015.
[3] European crude and age adjusted incidence by cancer, years of diagnosis 2000 and 2007: Analisys based on 83 population-based cancer registries. INFORMATION NETWORK ON RARE CANCERS IN EUROPE; 2015.
[4] Hashibe M, Brennan P, Benhamou S, Castellsague X, Chen C, Curado MP, et al. Alcohol drinking in never users of tobacco, cigarette smoking in never drinkers, and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. J Natl Cancer Inst 2007;99:777–89.
[5] Huang SH, Xu W, Waldron J, Siu L, Shen X, Tong L, et al. Refining American Joint Committee on Cancer/Union for International Cancer Control TNM stage and prognostic groups for human papillomavirus-related oropharyngeal carcinomas. J Clin Oncol 2015;33:836–45.
[6] Brana I, Siu LL. Locally advanced head and neck squamous cell cancer: treatment choice based on risk factors and optimizing drug prescription. Ann Oncol 2012;23(Suppl 10). x178-85.
[7] Leemans CR, Braakhuis BJ, Brakenhoff RH. The molecular biology of head and neck cancer. Nat Rev Cancer 2011;11:9–22.
[8] O'Sullivan B, Huang SH, Su J, Garden AS, Sturgis EM, Dahlstrom K, et al. Development and validation of a staging system for HPV-related oropharyngeal cancer by the International Collaboration on Oropharyngeal cancer Network for Staging (ICON-S): a multicentre cohort study. Lancet Oncol 2016;17:440–51.
[9] Tsao SW, Yip YL, Tsang CM, Pang PS, Lau VM, Zhang G, et al. Etiological factors of nasopharyngeal carcinoma. Oral Oncol 2014;50:330–8.
[10] de Bree R, Castelijns JA, Hoekstra OS, Leemans CR. Advances in imaging in the work-up of head and neck cancer patients. Oral Oncol 2009;45:930–5.
[11] Digonnet A, Hamoir M, Andry G, Haigentz Jr M, Takes RP, Silver CE, et al. Post-therapeutic surveillance strategies in head and neck squamous cell carcinoma. Eur Arch Otorhinolaryngol 2013;270:1569–80.
[12] Srinivasan A, Mohan S, Mukherji SK. Biologic imaging of head and neck cancer: the present and the future. AJNR Am J Neuroradiol 2012;33:586–94.
[13] Le Bihan D, Breton E, Lallemand D, Aubin ML, Vignaud J, Laval-Jeantet M. Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. Radiology 1988;168:497–505.
[14] Burdette JH, Elster AD, Ricci PE. Acute cerebral infarction: quantification of spin-density and T2 shine-through phenomena on diffusion-weighted MR images. Radiology 1999;212:333–9.

[15] Lemke A, Laun FB, Simon D, Stieltjes B, Schad LR. An in vivo verification of the intravoxel incoherent motion effect in diffusion-weighted imaging of the abdomen. Magn Reson Med 2010;64:1580–5.
[16] Driessen JP, van Kempen PM, van der Heijden GJ, Philippens ME, Pameijer FA, Stegeman I, et al. Diffusion-weighted imaging in head and neck squamous cell carcinomas: a systematic review. Head Neck 2015;37:440–8.
[17] Le Bihan D. Intravoxel incoherent motion imaging using steady-state free precession. Magn Reson Med 1988;7:346–51.
[18] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009;6:e1000100.
[19] Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529–36.
[20] Hayden JA, Cote P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. Ann Intern Med 2006;144:427–37.
[21] Hayden JA, van der Windt DA, Cartwright JL, Cote P, Bombardier C. Assessing bias in studies of prognostic factors. Ann Intern Med 2013;158:280–6.
[22] Dikaios N, Punwani S, Hamy V, Purpura P, Rice S, Forster M, et al. Noise estimation from averaged diffusion weighted images: can unbiased quantitative decay parameters assist cancer evaluation? Magn Reson Med 2014;71:2105–17.
[23] Ding Y, Hazle JD, Mohamed AS, Frank SJ, Hobbs BP, Colen RR, et al. Intravoxel incoherent motion imaging kinetics during chemoradiotherapy for human papillomavirus-associated squamous cell carcinoma of the oropharynx: preliminary results from a prospective pilot study. NMR Biomed 2015;28:1645–54.
[24] Guo W, Luo D, Lin M, Wu B, Li L, Zhao Y, et al. Pretreatment intra-voxel incoherent motion diffusion-weighted imaging (IVIM-DWI) in predicting induction chemotherapy response in locally advanced hypopharyngeal carcinoma. Med (Baltimore) 2016;95:e3039.
[25] Hauser T, Essig M, Jensen A, Gerigk L, Laun FB, Munter M, et al. Characterization and therapy monitoring of head and neck carcinomas using diffusion-imaging-based intravoxel incoherent motion parameters-preliminary results. Neuroradiology 2013;55:527–36.
[26] Hauser T, Essig M, Jensen A, Laun FB, Munter M, Maier-Hein KH, et al. Prediction of treatment response in head and neck carcinomas using IVIM-DWI: evaluation of lymph node metastasis. Eur J Radiol 2014;83:783–7.
[27] Lai V, Li X, Lee VH, Lam KO, Chan Q, Khong PL. Intravoxel incoherent motion MR imaging: comparison of diffusion and perfusion characteristics between nasopharyngeal carcinoma and post-chemoradiation fibrosis. Eur Radiol 2013;23:2793–801.
[28] Lai V, Li X, Lee VH, Lam KO, Fong DY, Huang B, et al. Nasopharyngeal carcinoma: comparison of diffusion and perfusion characteristics between different tumour stages using intravoxel incoherent motion MR imaging. Eur Radiol 2014;24:176–83.
[29] Lu Y, Jansen JF, Stambuk HE, Gupta G, Lee N, Gonen M, et al. Comparing primary tumors and metastatic nodes in head and neck cancer using intravoxel incoherent motion imaging: a preliminary experience. J Comput Assist Tomogr 2013;37:346–52.
[30] Marzi S, Piludu F, Vidiri A. Assessment of diffusion parameters by intravoxel incoherent motion MRI in head and neck squamous cell carcinoma. NMR Biomed 2013;26:1806–14.
[31] Sakamoto J, Imaizumi A, Sasaki Y, Kamio T, Wakoh M, Otonari-Yamamoto M, et al. Comparison of accuracy of intravoxel incoherent motion and apparent diffusion coefficient techniques for predicting malignancy of head and neck tumors using half-Fourier single-shot turbo spin-echo diffusion-weighted imaging. Magn Reson Imaging 2014;32:860–6.
[32] Sasaki M, Sumi M, Eida S, Katayama I, Hotokezaka Y, Nakamura T. Simple and reliable determination of intravoxel incoherent motion parameters for the differential diagnosis of head and neck tumors. PLoS ONE 2014;9:e112866.
[33] Sumi M, Nakamura T. Head and neck tumors: assessment of perfusion-related parameters and diffusion coefficients based on the intravoxel incoherent motion model. AJNR Am J Neuroradiol 2013;34:410–6.
[34] Sumi M, Van Cauteren M, Sumi T, Obara M, Ichikawa Y, Nakamura T. Salivary gland tumors: use of intravoxel incoherent motion MR imaging for assessment of diffusion and perfusion for the differentiation of benign from malignant tumors. Radiology 2012;263:770–7.
[35] Xiao Y, Pan J, Chen Y, Chen Y, He Z, Zheng X. Intravoxel incoherent motion-magnetic resonance imaging as an early predictor of treatment response to neoadjuvant chemotherapy in locoregionally advanced nasopharyngeal carcinoma. Med (Baltimore) 2015;94:e973.
[36] Xiao-ping Y, Jing H, Fei-ping L, Yin H, Qiang L, Lanlan W, et al. Intravoxel incoherent motion MRI for predicting early response to induction chemotherapy and chemoradiotherapy in patients with nasopharyngeal carcinoma. J Magn Reson Imaging 2016;43:1179–90.
[37] Yu XP, Hou J, Li FP, Wang H, Hu PS, Bi F, et al. Intravoxel incoherent motion diffusion weighted magnetic resonance imaging for differentiation between nasopharyngeal carcinoma and lymphoma at the primary site. J Comput Assist Tomogr 2016;40:413–8.
[38] Zhang SX, Jia QJ, Zhang ZP, Liang CH, Chen WB, Qiu QH, et al. Intravoxel incoherent motion MRI: emerging applications for nasopharyngeal carcinoma at the primary site. Eur Radiol 2014;24:1998–2004.

[39] Sumi M, Nakamura T. Head and neck tumours: combined MRI assessment based on IVIM and TIC analyses for the differentiation of tumors of different histological types. Eur Radiol 2014;24:223–31.

[40] de Bree R, van der Putten L, Brouwer J, Castelijns JA, Hoekstra OS, Leemans CR. Detection of locoregional recurrent head and neck cancer after (chemo) radiotherapy using modern imaging. Oral Oncol 2009;45:386–93.

[41] de Bree R, van der Putten L, van Tinteren H, Wedman J, Oyen WJ, Janssen LM, et al. Effectiveness of an (18)F-FDG-PET based strategy to optimize the diagnostic trajectory of suspected recurrent laryngeal carcinoma after radiotherapy: The RELAPS multicenter randomized trial. Radiother Oncol 2016;118:251–6.

[42] Lemke A, Laun FB, Simon D, Stieltjes B, Schad LR. An in vivo verification of the intravoxel incoherent motion effect in diffusion-weighted imaging of the abdomen. Magn Reson Med. 64:1580–5.

[43] Gasparini G, Weidner N, Maluta S, Pozza F, Boracchi P, Mezzetti M, et al. Intratumoral microvessel density and p53 protein: correlation with metastasis in head-and-neck squamous-cell carcinoma. Int J Cancer 1993;55:739–44.

[44] Gupta RK, Cloughesy TF, Sinha U, Garakian J, Lazareff J, Rubino G, et al. Relationships between choline magnetic resonance spectroscopy, apparent diffusion coefficient and quantitative histopathology in human glioma. J Neurooncol 2000;50:215–26.

[45] White ML, Zhang Y, Robinson RA. Evaluating tumors and tumorlike lesions of the nasal cavity, the paranasal sinuses, and the adjacent skull base with diffusion-weighted MRI. J Comput Assist Tomogr 2006;30:490–5.

[46] Lemke A, Stieltjes B, Schad LR, Laun FB. Toward an optimal distribution of b values for intravoxel incoherent motion imaging. Magn Reson Imaging 2011;29:766–76.

[47] Gurney-Champion OJ, Froeling M, Klaassen R, Runge JH, Bel A, van Laarhoven HW, et al. Minimizing the acquisition time for intravoxel incoherent motion magnetic resonance imaging acquisitions in the liver and pancreas. Invest Radiol 2016;51:211–20.

[48] Cui Y, Dyvorne H, Besa C, Cooper N, Taouli B. IVIM diffusion-weighted imaging of the liver at 3.0T: comparison with 1.5T. Eur J Radiol Open 2015;2:123–8.

[49] Kolff-Gart AS, Pouwels PJ, Noij DP, Ljumanovic R, Vandecaveye V, de Keyzer F, et al. Diffusion-weighted imaging of the head and neck in healthy subjects: reproducibility of ADC values in different MRI systems and repeat sessions. AJNR Am J Neuroradiol 2015;36:384–90.

[50] Le Bihan D, Breton E, Lallemand D, Grenier P, Cabanis E, Laval-Jeantet M. MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. Radiology 1986;161:401–7.

[51] Girometti R, Furlan A, Esposito G, Bazzocchi M, Como G, Soldano F, et al. Relevance of b-values in evaluating liver fibrosis: a study in healthy and cirrhotic subjects using two single-shot spin-echo echo-planar diffusion-weighted sequences. J Magn Reson Imaging 2008;28:411–9.

[52] Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 2009;45:228–47.