Journal of Cancer Research and Clinical Oncology (2019) 145:1811–1821
https://doi.org/10.1007/s00432-019-02934-y

REVIEW – CLINICAL ONCOLOGY

Accuracy of Raman spectroscopy in discrimination of nasopharyngeal carcinoma from normal samples: a systematic review and meta-analysis

Yuan Li[1] · Sihui Su[1] · Yingzhe Zhang[2] · Shiyao Liu[1] · Hongyu Jin[3] · Qianqing Zeng[1] · Lei Cheng[1]

Received: 10 January 2019 / Accepted: 10 May 2019 / Published online: 14 May 2019

Abstract

Objectives The aim of this review was to systematically evaluate the diagnostic accuracy of Raman spectroscopy (RS) in the identification of nasopharyngeal carcinomas from normal nasopharyngeal tissue.

Methods We searched six databases (PubMed, Embase, Cochrane Library, Web of Science, Scopus and CNKI) up to September 2018 for all published studies that assessed the diagnostic accuracy of RS in the detection of nasopharyngeal carcinomas. Non-qualifying studies were screened out in accordance with the specified exclusion criteria and relevant information about the diagnostic performance of RS extracted. A random effects model was adopted to calculate the pooled sensitivity, specificity, positive and negative likelihood ratios (PLR and NLR, respectively), diagnostic threshold and diagnostic odds ratio (DOR). Additionally, we conducted a summary receiver-operating characteristic (SROC) curve analysis and threshold analysis, reporting area under the curve (AUC) to evaluate the overall performance of RS.

Results Three studies examined RS analysis in vivo, the pooled sensitivity and specificity of RS of which were 0.90 and 0.91, respectively, with an AUC of 0.9617. Eighteen studies assessed ex vivo samples, for which RS exhibited particularly high accuracy for the analysis of blood plasma.

Conclusions RS was demonstrated to be a reliable technique for the detection of nasopharyngeal carcinoma with high accuracy, but additional studies are required to improve its performance and expand its application in ex vivo detection.

Keywords Nasopharyngeal Carcinoma · Raman spectroscopy · Diagnosis · Meta-analysis

Introduction

Nasopharyngeal carcinomas (NPCs) are among the most common malignant tumors of the head and neck. According to the latest statistics of global cancer in 2012, approximately 86,700 new cases of NPC and 50,800 deaths were reported Torre et al. (2015). The incidence of NPC has palpable geographical and gender characteristics. The incidence in males is 2–3 times higher than in females. Geographical differences are also noteworthy, with 92% of NPCs occurring in countries with lagging economic development (Torre et al. 2015; Huang et al. 2018). The region with the highest incidence is Southeast Asia, particularly in Malaysia, Indonesia, Singapore and Southern China (Chang and Adami 2006). An increasing number of clinicians have concluded that early diagnosis of NPC is essential for improving survival and quality of life. Currently, fiber nasopharyngeal laryngoscope, X-ray, CT, MRI in addition to serological examination are used in the diagnosis of NPC. Pathological

Yuan Li, Sihui Su, Yingzhe Zhang and Shiyao Liu contributed equally to this work.

✉ Lei Cheng
chengleidentist@163.com

[1] State Key Laboratory of Oral Diseases, National Clinical Research Center for Oral Diseases, West China Hospital of Stomatology, Sichuan University, 14 Renmin South Road Third Section, Chengdu 610041, China

[2] West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu 610041, China

[3] Department of Liver Surgery, Liver Transplantation Center, West China Hospital, Sichuan University, Chengdu 610041, China

sampling biopsy remains the gold standard for NPC diagnosis Wei and Sham (2005). However, these radiological and pathological techniques are invasive, time-consuming, lack sufficient accuracy and tend to be subjective. Thus, it would be of great benefit to introduce an advanced detection method that allowed accurate and rapid diagnosis.

In the last few years, many studies have described the use of Raman spectroscopy (RS) in the diagnosis of cancers, including those of the breast, stomach, skin, etc. (Deng et al. 2016; Ouyang et al. 2015; Zhang et al. 2018). RS evaluates a spectrum of scattered frequencies, which gives information about the chemical composition of functional groups by analyzing molecular vibrations (Lin et al. 2005). The principle of RS is based on the inelastic collision of photons with molecules, allowing the detection of the Stokes lines of the scattered photons. It currently plays an important role in diagnosis (Choo-Smith et al. 2002), due to its ability to detect material changes in the body during the early period of cancer by providing a spectral signature of the internal structure and conformation of cells (Das et al. 2017; Zhao 2018). Since it can analyze samples at the tissue level, it is noninvasive and so can help reduce the suffering and financial burden of traditional techniques. In addition, the analysis process is rapid, allowing real-time diagnostics. These characteristics contribute to RS having many advantages as a means of detection.

Many studies have examined the accuracy of RS in distinguishing differences between malignant tumors and normal tissue (Lin et al. 2017). In vitro and in vivo studies suggest that RS might be a valuable method for the identification of malignant tumor lesions of the nasopharynx. However, individual studies have been inconclusive due to insufficient sampling sizes and different experimental conditions and methodology, leading to contradictory results. In this review, we have aimed to systematically evaluate the literature to determine the diagnostic accuracy of RS to discriminate nasopharyngeal carcinoma and normal nasopharyngeal tissue.

Materials and methods

Literature search

Published articles were systematically searched from six databases (PubMed, Embase, Cochrane Library, Web of Science, Scopus and CNKI) up to September 2018, without any limit to start date. The search keywords were: "Raman" and "Nasopharyngeal neoplasm". The keyword "Raman" was also expanded to raman spectrum analysis, raman spectroscopy, raman spectrum, raman optical activity, raman scattering, raman spectra, raman spectroscope and raman spectrometry. For the keyword "Nasopharyngeal", the terms rhinopharynxes, rhinopharynges, rhinopharynx, nasopharynges, nasopharynxes and choanae were also searched, while neoplasia, neoplasias, neoplasm, tumor, cancer, malignant neoplasm and benign neoplasm were searched as expansions of "neoplasm". The search strategy is displayed in Table 1. Related articles and the reference lists of the literature identified in the search were searched to identify all relevant studies, abstracts and citations. The full texts of all potential studies were downloaded to ascertain whether they were eligible for inclusion.

Study selection criteria

Studies were included in the review if they complied with the following inclusion criteria: (1) nasopharyngeal carcinoma samples were detected by RS and diagnosed using histopathology as the gold reference standard; (2) a control group was included without nasopharyngeal carcinoma; (3) sufficient data were presented to construct a 2×2 table including true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) and (4) studies possessed a low risk of bias with a quality assessment score ≥ 10.

Exclusion criteria: (1) studies involving nonhuman subjects; (2) studies that had no control group or were case reports or a case series; (3) duplicate reports or reviews; (4) studies with a high risk of bias.

Data extraction and quality assessment

The following data were extracted independently by two experienced investigators from each study: the first author's name, geographical location, number of patients, samples and spectra, mean age of subjects, sample type, methodological and technical data such as diagnostic algorithm, study design, RS technique, TP, TN, FP and FN. The quality of each study was assessed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool Whiting et al. (2003), an established, evidence-based tool for systematic reviews of diagnostic studies, designed for diagnostic accuracy. The number of 'Yes' scores in each study was defined as the quality score. According to the selection criteria, studies scoring at least 10 were deemed to have a low risk of bias and high quality; otherwise, studies scoring less than 10 were deemed to have a high risk of bias and low quality.

Statistical analysis

To assess the diagnostic accuracy of Raman spectra for nasopharyngeal carcinoma, we calculated the pooled sensitivity, specificity, positive likelihood ratios (PLR), negative likelihood ratios (NLR), diagnostic threshold and diagnostic odds ratio (DOR), each with 95% confidence intervals (CI). In order to consider variation among the studies, we selected a

Table 1 Search strategies used in this review

Databases	Steps	Strategies
PubMed	#1	"Nasopharynx"[Mesh] OR Rhinopharynxes OR Rhinopharynges OR Rhinopharynx OR Nasopharynges OR Nasopharynxes OR Choanae
	#2	"Neoplasms"[Mesh] OR Neoplasms OR Neoplasia OR Tumor OR Neoplasm OR Cancer OR Malignant Neoplasm OR Benign Neoplasm
	#3	(#1 AND #2) OR "Nasopharyngeal Neoplasms"[Mesh]
	#4	Raman Spectrum Analysis OR Raman Spectroscopy OR Spectroscopy, Raman OR Analysis, Raman Spectrum OR Raman Optical Activity Spectroscopy OR Raman Scattering OR Scattering, Raman OR Raman spectra OR Raman spectrum OR Raman spectroscope OR Raman spectrometry OR "Spectrum Analysis, Raman"[Mesh]
	#5	#3 AND #4
Embase	#1	'neoplasm'/exp OR neoplasia' OR 'tumor' OR 'cancer' OR 'malignant neoplasm' OR 'benign neoplasm'
	#2	'rhinopharynx'/exp OR 'rhinopharynxes' OR 'rhinopharynges' OR'rhinopharynx' OR'nasopharynges' OR 'nasopharynxes' OR 'choanae'
	#3	(#1 AND #2) OR 'nasopharynx neoplasms'/exp
	#4	'raman spectrum analysis' OR 'raman spectroscopy' OR'raman optical activity spectroscopy' OR 'raman scattering'OR 'raman spectra' OR 'raman spectrum' OR 'raman spectroscope' OR 'raman spectrometry'
	#5	#3 AND #4
Cochrane library	#1	MeSH descriptor: [Spectrum Analysis, Raman] explode all trees
	#2	MeSH descriptor: [Nasopharyngeal Neoplasms] explode all trees
	#3	#1 AND #2
CNKI and Scopus	#1	Raman
	#2	Nasopharyngeal Neoplasms
	#3	#1 AND #2
Web of Science	#1	Nasopharynx OR Rhinopharynxes OR Rhinopharynges OR Rhinopharynx OR Nasopharynges OR Nasopharynxes OR Choanae
	#2	Neoplasms OR Neoplasms OR Neoplasia OR Tumor OR Neoplasm OR Cancer OR Malignant Neoplasm OR Benign Neoplasm
	#3	(#1 AND #2) OR Nasopharyngeal Neoplasms
		Raman Spectrum Analysis OR Raman Spectroscopy OR Raman Optical Activity Spectroscopy OR Raman Scattering OR Raman spectra OR Raman
	#4	spectrum OR Raman spectroscope OR Raman spectrometry
	#5	#3 AND #4

fixed effects model (Mantel–Haenszel method) or a random effects model (DerSimonian–Laird method) to pool those parameters. Fixed effects models assume that all studies in a meta-analysis are drawn from a common population. In our meta-analysis, we pooled those parameters by adopting a random effects model (DerSimonian–Laird method), making the assumption that the studies were drawn from populations that differed from each other in a way that could affect the final result (Lean et al. 2009; Melsen et al. 2014).

In order to investigate the effect of thresholds on consequences, a summary receiver-operating characteristic (SROC) curve and threshold analyses were conducted, with area under the curve (AUC) calculated to evaluate the overall performance of RS. The SROC curves were not shoulder-like, indicating that thresholds may have no impact on the results. An excellent diagnostic effect is demonstrated when an AUC value is between 0.9 and 1, good for an AUC value of 0.8–0.9; fair for an AUC value of 0.7–0.8 and poor when the AUC value is 0.6–0.7. The diagnostic method can be deemed to have failed when the AUC is between 0.5 and 0.6 (Metz 1978). To further assess the heterogeneity of the studies, a DerSimonian–Laird test (Q statistic) and the inconsistency index (I^2) statistic were calculated. In this way, the Q statistic identified the presence or absence of heterogeneity and the I^2 index classified the degree of heterogeneity (Huedo-Medina et al. 2006), which was considered significant for an I^2 index > 50% and a P value < 0.05 (Higgins et al. 2003). Publication bias was assessed using a Deeks funnel plot asymmetry test, which was conducted using Stata 12.0 software. Publication bias is considered to be present when the P value < 0.05 (Begg and Mazumdar 1994).

All data analyses above were performed by using MetaDiSc 1.4 and Stata 12.0 software. For all tests, P values less than 0.05 were considered statistically significant.

Results

Study selection

The initial literature search yielded 373 potential articles (from PubMed, EmBase, Cochrane, Web of Science, Scopus and CNKI) in which 285 articles were selected after duplicates were removed. Two hundred and forty-four studies that were not relevant were excluded, so that 41 potentially relevant articles remained after reading the article abstracts and titles. Of these, 18 articles were excluded after studying the full-text because they were reviews or because of lack of relevance. Ultimately, 21 studies (Chen 2010; Feng 2011; Feng et al. 2010, 2017; Li 2012, 2017; Li et al. 2013, 2015; Lin et al. 2014, 2017, 2018a, b; Lu 2014; Ming et al. 2017; Pan et al. 2011; Qiu 2016; Qiu et al. 2016a, b; Sun et al. 2016; Wu et al. 2018; Wu 2015) were selected as eligible and included in the meta-analysis following assessment of quality, in accordance with the selection criteria, because two articles (Lin et al. 2017; Li et al. 2012) were assessed as low quality and thus excluded. No additional articles were found to be relevant from the reference lists of the selected publications. The study screening and selection process and metrics are shown in Fig. 1.

Description of studies included in the review

Among the 21 eligible studies, 9 (Chen 2010; Feng 2011; Li 2012, 2017; Lin et al. 2018; Lu 2014; Pan et al. 2011; Qiu 2016; Wu 2015) were published in Chinese, whereas the other 12 (Lin et al. 2014, 2017, 2018; Feng et al. 2010, 2017; Li et al. 2013, 2015; Ming et al. 2017; Qiu et al. 2016a, b; Sun et al. 2016; Wu et al. 2018) were published in English. All studies were conducted by research teams in China except one (Ming et al. 2017), which was conducted in Singapore. Histopathological samples were utilized as the gold standard for all studies. The mean age in all studies included in the review ranged from 42 to 55 years. Three studies (Lin et al. 2017, 2018; Ming et al. 2017) evaluated in vivo tissue, while the others assessed samples ex vivo, the sample types of which were urine in one study (Feng et al. 2017), tissue DNA in two (Li 2012; 2017), saliva in three (Lin et al. 2018; Qiu 2016; Qiu et al. 2016), blood plasma in seven (Chen 2010; Feng 2011; Feng et al. 2010; Lin et al. 2014; Lu 2014; Pan et al. 2011; Wu et al. 2018) and nasal tissue in five (Li 2012; Li et al. 2013, 2015; Sun et al. 2016; Wu 2015). Various diagnostic algorithms were utilized to analyze the result of RS, including principal components analysis (PCA)–linear discriminant analysis (LDA) ($n=17$) (Lin et al. 2014, 2017, 2018a, b; Feng 2011; Feng et al. 2010; Li 2012; Li et al. 2013; Lu 2014; Ming et al. 2017; Pan et al. 2011; Qiu 2016; Qiu et al. 2016; Sun et al. 2016; Wu et al. 2018; Wu 2015), PCA–discriminant function analysis (DFA) ($n=3$) (Chen 2010; Li 2017; Qiu et al. 2016) and partial least squares (PLS)–linear discriminant analysis (LDA) ($n=2$) (Feng et al. 2017; Li et al. 2015). The Raman spectra were obtained using three forms of RS, including surface-enhanced Raman spectroscopy (SERS) ($n=12$) (Feng 2011; Feng et al. 2010, 2017; Li 2017; Lin et al. 2014, 2018; Lu 2014; Qiu 2016; Qiu et al. 2016a, b; Wu et al. 2018; Wu 2015), confocal micro-Raman spectroscopy (CMRS) ($n=7$) (Lin et al. 2017, 2018; Li 2012; Ming et al. 2017) and fiber-optic Raman spectroscopy (FORS) ($n=4$) (Lin et al. 2017, 2018; Li 2012; Lu 2014). Detailed information about each study is shown in Table 2.

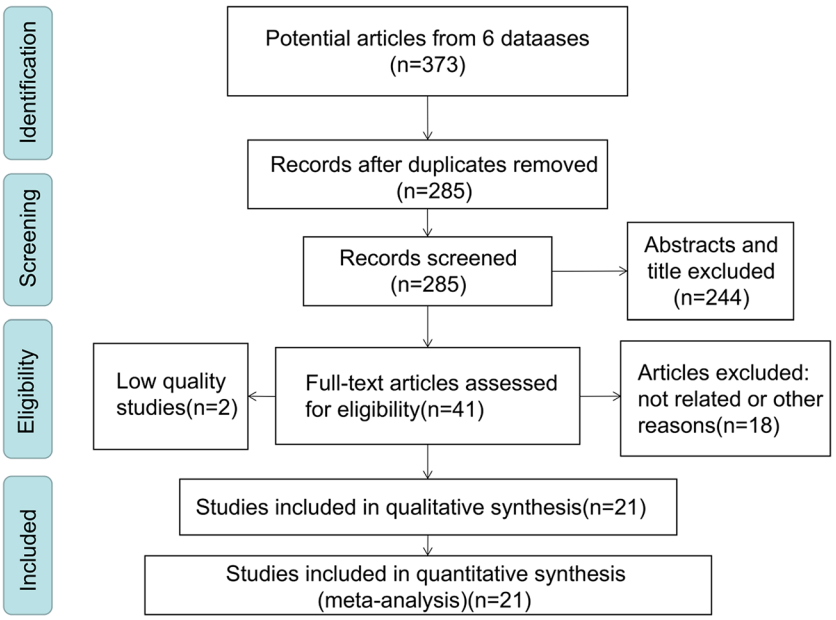

Fig. 1 Systematic review flow diagram

Table 2 General information of the studies included in the review

Reference	Country	*N*1	*N*2	*N*3	Mean age	Sample type	Diagnostic algorithm	TP	FP	FN	TN	Raman spectroscopy
Chen (2010)	C	46	46	230	U	EV: blood plasma	PCA–DFA	29	0	0	17	CMRS
Feng et al. (2010)	C	76	U	U	42	EV: blood plasma	PCA–LDA	39	0	4	33	SERS
Feng (2011)	C	76	76	228	42	EV: blood plasma	PCA–LDA	43	0	3	30	SERS
Feng (2017)	C	114	114	114	55	EV: urine	PLS–LDA	58	0	4	52	SERS
Li (2017)	C	115	115	230	50	EV: DNA	PCA–DFA	65	7	13	30	SERS
Li (2012a)	C	74	74	U	47	EV: tissue	PCA–LDA	48	4	4	18	CMRS
Li (2012)	C	48	48	265	43.2	EV: tissue	PCA–LDA	28	3	4	13	FORS
Li (2013)	C	74	U	U	46.5	EV: tissue	PCA–LDA	48	4	4	18	CMRS
Li (2015a)	C	48	48	U	U	EV: tissue	PCA–LDA	26	2	6	14	CMRS
Li (2015b)	C	48	48	U	U	EV: tissue	PLS–LDA	27	2	5	14	CMRS
Lin (2014a)	C	85	85	U	U	EV: Blood plasma	PCA–LDA	21	10	4	50	SERS
Lin (2014b)	C	135	135	U	U	EV: blood plasma	PCA–LDA	69	3	6	57	SERS
Lin (2018)	C	60	283	283	53.5	IV	PCA–LDA	126	11	15	131	FORS
Lin (2017a)	C	95	204	3731	52	IV	PCA–LDA	102	7	7	88	FORS
Lin (2017b)	C	95	204	3731	52	IV	PCA–LDA	98	11	11	84	FORS
Lin (2017c	C	95	204	3731	52	IV	PCA–LDA	97	9	12	86	FORS
Lin (2018)	C	94	94	208	48	EV: Saliva	PCA–LDA	43	2	3	46	SERS
Lu (2014)	C	44	44	132	U	EV: blood plasma	PCA–LDA	20	1	2	21	SERS
Ming (2017)	Singapore	42	64	64	53.9	IV	PCA–LDA	20	2	2	40	FORS
Pan (2011)	C	65	65	325	U	EV: blood plasma	PCA–LDA	26	3	3	33	CMRS
Qiu (2016a)	C	112	112	U	U	EV: saliva	PCA–LDA	43	5	9	55	SERS
Qiu (2016b)	C	91	91	U	47.9	EV: DNA	PCA–DFA	55	5	6	25	SERS
Qiu (2016c)	C	62	62	186	U	EV: saliva	PCA–LDA	26	4	6	26	SERS
Sun (2016)	C	74	74	U	45	EV: tissue smear	PCA–LDA	34	5	5	30	CMRS
Wu (2018)	C	70	70	U	U	EV: blood plasma	PCA–LDA	38	2	2	28	SERS
Wu (2015a)	C	74	74	370	45	EV: tissue smear	PCA–LDA	32	6	7	29	CMRS
Wu (2015b)	C	67	67	U	45	EV: tissue smear	PCA–LDA	27	6	7	27	SERS
Wu (2015c)	C	78	78	U	U	EV: tissue	PCA–LDA	35	3	6	34	CMRS

C China, *U* unknown, *N1* number of patients, *N2* number of samples, *N3* number of spectra, *EV* ex vivo, *IV* in vivo, *PCA* principal component analysis, *LDA* linear discriminate analysis, *PLS* partial least squares, *DFA* discriminant function analysis, *(A)(B)(C)* a same author published articles in a same year, *(a)(b)(c)* there were several subgroup studies in one article, *SERS* surface-enhanced Raman spectroscopy, *CMRS* confocal micro-Raman spectroscopy, *FORS* fiber-optic Raman spectroscopy

Assessment of study quality

Two reviewers independently evaluated the risk of bias and quality for the initial 23 enrolled studies (Chen 2010; Feng et al. 2010, 2017; Feng 2011; Pan et al. 2011; Li et al. 2012; Li 2012, 2017; Li et al. 2013, 2015; Lin et al. 2014, 2017, 2018; Lu 2014; Ming et al. 2017; Qiu 2016; Qiu et al. 2016a, b; Sun et al. 2016; Wu et al. 2018; Wu 2015) according to QUADAS guidenlines. All QUADAS items were used to evaluate eligible articles. The results of the evaluation of each study are provided in Table 3. Of the 23 studies, 9 had a total quality score of 11 (78.6% YES rate) and 12 had a score of 10 (71.4% YES rate). However, 2 studies (Li et al. 2012; Lin et al. 2017) scored only 9 (64.3% YES rate) and 7 (50% YES rate), respectively, indicating that more than 5 questions were assessed as NO. Such scoring suggests that they exhibited high risk of bias and would reduce the quality of the meta-analysis and accordingly were excluded from the final meta-analysis through lack of compliance with the selection criteria, their total quality being significantly lower than the others. Finally, 21 studies were included in the meta-analysis.

Pooled results

In vivo group

Three studies (Lin et al. 2017, 2018; Ming et al. 2017) examined tissue in vivo and the pooled sensitivity and specificity of RS in this group (in vivo group) were 0.90

Table 3 Quality assessment of the included studies using the QUADAS tool

Reference	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Score	Rate of Y
Chen (2010	Y	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	U	U	11	78.6
Feng (2010)	Y	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	U	U	11	78.6
Feng (2011)	Y	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	U	U	11	78.6
Feng (2017)	U	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	U	U	10	71.4
Li (2017)	Y	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	U	U	11	78.6
Li (2012)	Y	U	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	U	U	10	71.4
Li (2013)	Y	N	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	U	U	10	71.4
Li (2015)	U	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	U	U	10	71.4
Li (2017)	U	N	Y	Y	Y	Y	Y	Y	Y	U	U	U	U	U	7	50
Lin (2014)	Y	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	U	U	U	10	71.4
Lin (2018)	Y	N	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	U	U	10	71.4
Lin (2017)	Y	N	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	U	U	10	71.4
Lin (2015)	Y	Y	Y	U	Y	Y	Y	Y	Y	U	Y	U	U	U	9	64.3
Lin (2018)	Y	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	U	U	11	78.6
Lu (2014)	Y	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	U	U	U	10	71.4
Ming (2017)	Y	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	U	U	11	78.6
Pan (2011)	Y	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	U	U	U	10	71.4
Qiu (2016a)	Y	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	U	U	U	10	71.4
Qiu (2016b)	Y	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	U	U	11	78.6
Qiu (2016c)	Y	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	U	U	11	78.6
Sun (2016)	Y	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	U	U	11	78.6
Wu (2018)	Y	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	U	U	U	10	71.4
Wu (2015)	Y	Y	Y	Y	Y	Y	Y	Y	Y	U	U	Y	U	U	10	71.4

Q1. Was the spectrum of patients representative of the patients who will receive the test in practice? Q2. Were selection criteria clearly described? Q3. Is the reference standard likely to correctly classify the target condition? Q4. Is the time period between reference standard and index test short enough to be reasonable? Q5. Did the whole sample, or a random selection of the sample, receive verification using a reference standard of diagnosis? Q6. Did patients receive the same reference standard regardless of the index test result? Q7. Was the reference standard independent of the index test (i.e., the index test did not form part of the reference standard)? Q8. Was the execution of the index test described in sufficient detail to permit replication of the test? Q9. Was the execution of the reference standard described in sufficient detail to permit its replication? Q10. Were the index test results interpreted without knowledge of the results of the reference test? Q11. Were the reference standard results interpreted without knowledge of the results of the index test? Q12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice? Q13. Were interpretable/intermediate test results reported?Q14. Were withdrawals from the study explained?

Y yes, *N* no, *U* unclear, *QUADAS* Quality assessment of diagnostic accuracy studied

(95% CI 0.87–0.93, $P = 0.761$, $I^2 = 0.0\%$) and 0.91 (95% CI 0.89–0.94, $P = 0.678$, $I^2 = 0.0\%$), respectively (Fig. 2). Pooled PLR and NLR were 10.23 (95% CI 7.62–13.80) and 0.11 (95% CI 0.08–0.14), respectively. The DOR was 97.65 (95% CI 62.64–152.23). An SROC curve was plotted to summarize overall diagnostic accuracy, for which the AUC was 0.9617.

Tissue/ex vivo group

Five studies (Li 2012; Li et al. 2013, 2015; Wu 2015; Sun et al. 2016) examined tissue ex vivo. Their pooled sensitivity and specificity of Raman spectroscopy to discriminate were 0.86 (95% CI 0.82–0.90) and 0.85 (95% CI 0.80–0.89), respectively. Plots are shown in Fig. 3. The pooled PLR and NLR were 5.44 (95% CI 4.00–7.40) and 0.17 (95% CI 0.13–0.22), respectively. The DOR was 34.22 (95% CI 21.17–55.32). An SROC curve was plotted to evaluate overall diagnostic accuracy, with an AUC of 0.9193.

Blood plasma/ex vivo group

Seven studies (Chen 2010; Feng et al. 2010; Feng 2011; Pan et al. 2011; Lin et al. 2014; Lu 2014; Wu et al. 2018) examined blood plasma samples ex vivo. Their pooled sensitivity and specificity for discrimination using Raman spectroscopy were 0.92 (95% CI 0.89–0.95) and 0.93 (95% CI 0.90–0.96), respectively. Data are also plotted in Fig. 3. The pooled PLR and NLR were 14.69 (95% CI 6.69–32.27) and 0.10 (95% CI 0.07–0.66), respectively. The DOR was 165.74 (95% CI 64.53–425.6), also

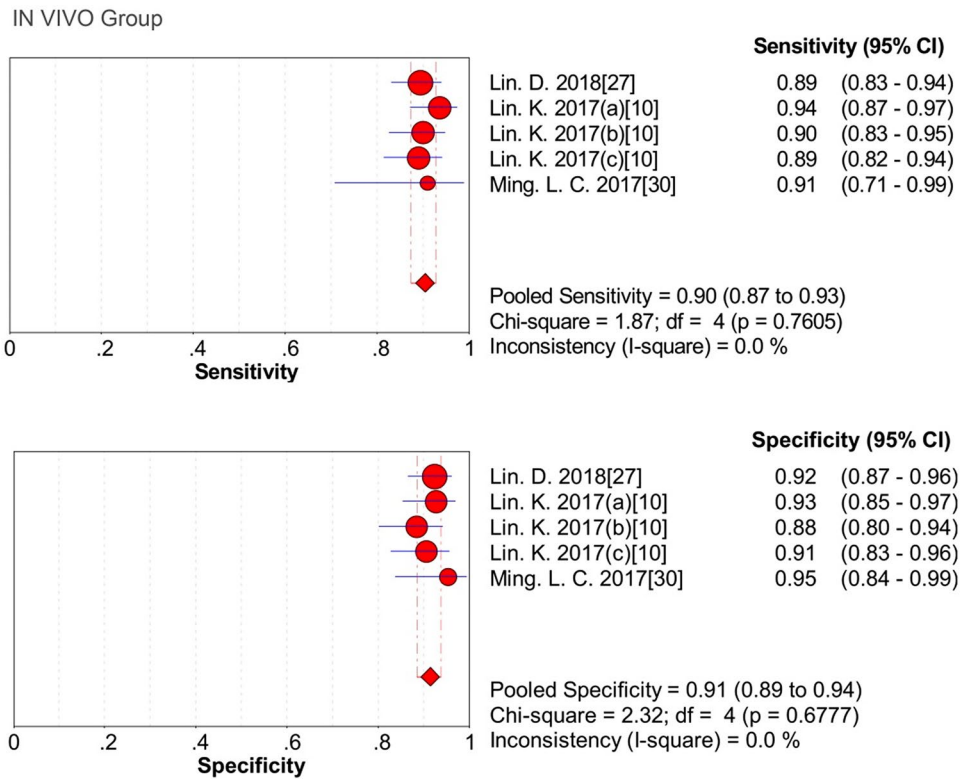

Fig. 2 Individual study and pooled estimates of sensitivity, specificity and their 95% confidence intervals (CIs) of RS in the in vivo group

demonstrating very high accuracy. An SROC curve was plotted to evaluate overall diagnostic accuracy, with an AUC of 0.9673.

Saliva/ex vivo group

Three studies (Qiu 2016; Qiu et al. 2016; Lin et al. 2018) examined saliva samples ex vivo. Their pooled sensitivity and specificity of Raman spectroscopy regarding discrimination were 0.86 (95% CI 0.79–0.92) and 0.92 (95% CI 0.86–0.96), respectively, with data shown in Fig. 3. The pooled PLR and NLR were 9.76 (95% CI 5.04–18.88) and 0.16 (95% CI 0.09–0.29), respectively. The DOR was 68.28 (95% CI 19.69–236.81). An SROC curve, summarizing overall diagnostic accuracy, yielded an AUC of 0.9614.

DNA/ex vivo group

Two studies (Qiu et al. 2016; Li 2017) examined DNA samples ex vivo, with pooled sensitivity and specificity of discrimination by RS being 0.86 (95% CI 0.80–0.92) and 0.82 (95% CI 0.71–0.90), respectively. Plots of the data are shown in Fig. 3. The pooled PLR and NLR were 4.80 (95% CI 2.86–8.04) and 0.17 (95% CI 0.10–0.29), respectively. The DOR was 28.77 (95% CI 12.99–63.70).

Urine/ex vivo group

Only one study (Feng et al. 2017) examined urine samples ex vivo and so meta-analysis could not be performed. The sensitivity and specificity for this study were 0.935 and 1, respectively.

Publication bias and heterogeneity

The results of the evaluation of publication bias and heterogeneity in each group are shown in Table 4. A Deeks funnel plot asymmetry test indicated that no significant publication bias was found in any group.

Discussion

This meta-analysis was conducted to assess for the first time the accuracy of Raman spectroscopy in differentiating nasopharyngeal carcinoma from healthy biological samples. A total of 21 high-quality publications were selected, with no restriction to language. As a consequence, except for 12 English publications, an additional 9 Chinese publications were found. As a statistical study, it was probably not appropriate to include studies published in any language other than English, for damaging the reproducibility of readers whose native language was not Chinese. But it was reasonable

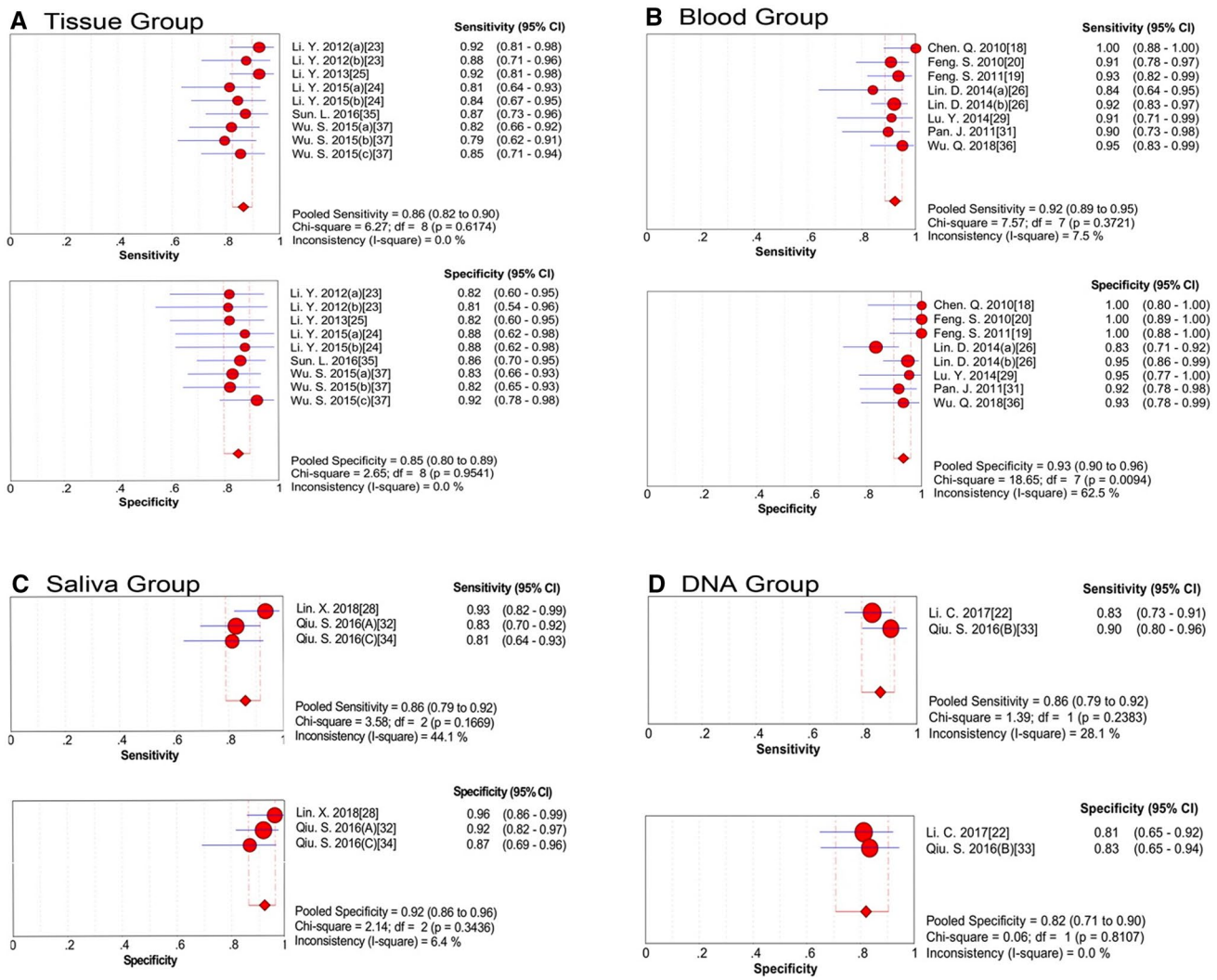

Fig. 3 Individual study and pooled estimates of sensitivity, specificity and their 95% confidence intervals (CIs) of RS in the ex vivo groups

in our study. Although NPC is a global disease, it is more severe in southern China than in Europe or America. The incidence and mortality of NPC in China are 3.09/100,000 and 1.57/100,000, respectively, while the global age-standardized incidence and mortality are 2.17/100,000 and 1.08/100,000, respectively (Wei et al. 2017). To control for this, Chinese researchers have conducted extensive studies on the application of Raman spectroscopy in the diagnosis of NPC. Even those English publications selected for our review were mostly conducted by Chinese teams. So the Chinese publications were as important for our analysis as the English publications. Therefore, we chose to select them to ensure the evidence was comprehensive and globally relevant, which is important for a systematic review and meta-analysis.

As can be seen in Table 4, a number of important results were observed. Firstly, the diagnostic performance of RS analysis of NPC tissue in vivo was better than tissue ex vivo, with AUC values of 0.9617 and 0.9193, respectively. The reason for this may be that ex vivo nasopharyngeal tissues may suffer desiccation following tissue explantation (Bergholt et al. 2013). Besides, devices used in the three in vivo studies were modified fiber-optic Raman spectroscopy systems, which have improved accuracy compared to bulky commercial micro-Raman spectrograph systems used in ex vivo studies. Secondly, among the ex vivo samples, saliva (AUC = 0.9673) and blood plasma (AUC = 0.9641) performed better than ex vivo tissue. The principal components of blood plasma and saliva are derived mostly from the secretion or metabolites of cells. Changes to various components of body fluid can directly reflect physiological and pathological changes due to cancer cells in vivo (Gahan et al. 2008), thus affecting the Raman spectrum. The combination of Raman spectrum analysis of blood plasma or

Table 4 Pooled estimation of sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, diagnostic odds ratio and area under the curve for Raman spectroscopy

Groups (*N*)	SEN (95% CI *P*, I^2)	SPE (95% CI *P*, I^2)	PLR (95% CI *P*, I^2)	NLR (95% CI *P*, I^2)	DOR (95% CI *P*, I^2)	AUC	Publication bias(*P**)
In vivo (3)	0.90(0.87–0.93, 0.760, 0.0%)	0.915(0.87–0.94, 0.680, 0.0%)	10.23(7.62–13.80, 0.670, 0.0%)	0.11(0.08–0.14, 0.770, 0.0%)	97.65(62.64–152.23, 0.620, 0.0%)	0.9617	0.5
Ex vivo (18)							
Tissue (5)	0.86(0.82–0.90, 0.617, 0.0%)	0.85(0.80–0.89, 0.954, 0.0%)	5.44(4.00–7.40, 0.97, 0.0%)	0.17(0.13–0.22,0.73, 0.0%)	34.22(21.17–55.32, 0.91, 0.0%)	0.9193	0.74
Blood plasma (7)	0.92(0.89–0.95, 0.372, 7.5%)	0.93(0.90–0.96, 0.009, 62.5%)	14.69(6.69–32.27, 0.019, 58.1%)	0.10(0.07–0.14, 0.66, 0.0%)	165.74(64.53–425.66, 0.11, 39.9%)	0.9673	0.81
Saliva (3)	0.86(0.79–0.92, 0.167, 44.1%)	0.92(0.86–0.96, 0.344, 6.4%)	9.76(5.04–18.88, 0.273, 22.9%)	0.16(0.09–0.29, 0.174, 42.9%)	68.28(19.69–236.81, 0.104, 55.7%)	0.9641	0.77
DNA (2)	0.86(0.80–0.92, 0.238, 28.1%)	0.82(0.71–0.90, 0.811, 0.0%)	4.80(2.86–8.04, 0.701, 0.0%)	0.17(0.10–0.29, 0.238, 28.0%)	28.77(12.99–63.70, 0.361, 0.0%)	_	_
Urine (1)	0.935	1	_	_	_	_	_

AUC area under curve, *P P* value of the Q statistic, I^2 the inconsistency index, *CI* confidence interval, *DOR* diagnostic odds ratio, *NLP* negative likelihood ratio, *PLR* positive likelihood ratio, *SEN* sensitivity, *SPE* specificity, *N* number of studies, *P* P* value of the Deeks funnel plot asymmetry test

saliva and SERS technology has been increasingly used for the diagnosis of NPC. SERS can remedy for the inherent defect of weak Raman spectrum signal by not only providing molecule specificity but also adding high sensitivity. The enhancement factor can be 10^6–10^{10} at most when the Ag or Au nanoparticles are mixed with analyte molecules (Bell and Sirimuthu 2008). Therefore, blood plasma and saliva have greater accuracy than ex vivo tissue because the tissues were not analyzed using SERS. Thirdly, DNA (sensitivity = 0.82, specificity = 0.86) performed the worst of all sample types. This may be because body fluids and tissues can provide comprehensive information about the molecular characteristics of nucleic acids, proteins and other molecules in cancer, but DNA can only provide a part of that information. Fourthly, urine (sensitivity = 0.935, specificity = 1) has the potential to be an equally good choice compared with saliva and blood plasma, as it is an important biological fluid for assessment of a patient's physical state and contains general tumor markers for different types of cancer (Ma et al. 2004). However, we found only one relevant research study and so additional research is required to establish its accuracy.

RS can be considered a promising diagnostic tool for nasopharyngeal carcinoma with several advantages. Firstly, RS is a noninvasive technique that does no harm to patients when used ex vivo. It also does not damage the surface of the nasopharyngeal tissue or invade it when the device is inserted into the nasopharynx. Secondly, RS can easily be performed for collection of spectral information, which requires no special staining or preparation (Fox et al. 2014), making real-time diagnosis possible and the ability to avoid surgical workflow disruption. All that is required is aiming the light source at the sample or lesion tissue through a micro-lens or fiber-optic probe. The handheld Raman probe is relatively small and easy to operate during the procedure (Molckovsky et al. 2003). Thirdly, RS only takes a few minutes to conduct to obtain accurate spectroscopic data, while traditional analytic techniques, such as histopathology requires several hours or days to prepare samples and obtain diagnostic results (Huang et al. 2003). Although the data analysis procedure is time-consuming and complicated for RS, it is still shorter for a well-trained professional. Fourthly, the excellent accuracy of RS in diagnosis will contribute to a decrease in demand for expensive tests to guarantee the correct diagnosis (Leslie et al. 2012). However, there are still a few drawbacks that should be taken into consideration. The analysis of data collected by RS requires extensive skill, which requires the training of qualified algorithm designers (Zhao et al. 2015). Although the accuracy of RS is comparatively high, it is not as flexible and accurate as traditional tests and there is no evidence that RS has really improved diagnostic efficiency or prognostic capability.

In conclusion, RS is an accurate, noninvasive, convenient, rapid and inexpensive technique for the real-time and early diagnosis of NPC. It has great potential for clinical application according to current research progress, especially when applied to blood plasma or saliva in vitro and nasopharyngeal tissue in vivo. But RS may not yet be ready to replace the gold standard or other auxiliary examinations for the eventual diagnosis of NPC. Additional research is required to validate RS for clinical application. Additional in vitro

and in vivo studies with larger sampling sizes are required. It is expected that the accuracy of RS will be improved until it is comparable with existing techniques, using more creative methods, such as cellulose template techniques (Lu 2014) and plasma electrophoresis (Feng 2011) reported in earlier studies, which has increased the accuracy of RS in a number of ways. Also, laboratory RS devices used in existing studies should be further refined for clinical use.

Our review also had several limitations. Firstly, the number of samples in each research study was small. More studies with larger numbers of patients are required. Secondly, in some subgroups such as the DNA and urine groups, the number of studies was grossly inadequate. Thirdly, multiple algorithms, and different RS techniques were used in the studies in this review, which interfered with comparisons between groups. Fourthly, we were not able to assess the accuracy of RS for distinguishing early and advanced nasopharyngeal neoplasms owing to the lack of relevant studies.

Conclusions

RS was demonstrated to be a reliable approach for the diagnosis of nasopharyngeal carcinoma with high accuracy. Meanwhile, RS can be utilized to analyze both body fluids ex vivo and in vivo tissue. It has huge potential for use as a technique in clinical application, but several weaknesses remain to be addressed. Therefore, more studies should be conducted to improve the performance of RS and generalize its use in real-time and early NPC diagnosis.

Compliance with ethical standards

Conflict of interest The authors declare that there are no potential conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Begg CB, Mazumdar M (1994) Operating characteristics of a rank correlation test for publication bias. Biometrics 50(4):1088–1101

Bell SE, Sirimuthu NM (2008) Quantitative surface-enhanced Raman spectroscopy. Chem Soc Rev 37(5):1012–1024

Bergholt MS et al (2013) Multivariate reference technique for quantitative analysis of fiber-optic tissue Raman spectroscopy. Anal Chem 85(23):11297–11303

Chang ET, Adami HO (2006) The enigmatic epidemiology of nasopharyngeal carcinoma. Cancer Epidemiol Biomarkers Prev 15(10):1765–1777

Chen Q (2010) Research on raman spectra of serum in nasopharyngeal carcinoma. Fujian Medical University, Fuzhou

Choo-Smith LP et al (2002) Medical applications of Raman spectroscopy: from proof of principle to clinical implementation. Biopolymers 67(1):1–9

Das NK et al (2017) Raman plus X: biomedical applications of multimodal raman spectroscopy. Sensors (Basel) 17(7):1592

Deng K et al (2016) Rapid discrimination of malignant breast lesions from normal tissues utilizing raman spectroscopy system: a systematic review and meta-analysis of in vitro studies. PLoS One 11(7):e0159860

Feng S (2011) Study on detection methods of nasopharyngeal and gastric cancer based on surface enhanced Raman spectroscopy. Fujian Normal University, Fuzhou

Feng S et al (2010) Nasopharyngeal cancer detection based on blood plasma surface-enhanced Raman spectroscopy and multivariate analysis. Biosens Bioelectron 25(11):2414–2419

Feng S et al (2017) A noninvasive cancer detection strategy based on gold nanoparticle surface-enhanced raman spectroscopy of urinary modified nucleosides isolated by affinity chromatography. Biosens Bioelectron 91:616–622

Fox SA et al (2014) Raman spectroscopy differentiates squamous cell carcinoma (SCC) from normal skin following treatment with a high-powered CO2 laser. Lasers Surg Med 46(10):757–772

Gahan PB, Anker P, Stroun M (2008) Metabolic DNA as the origin of spontaneously released DNA? Ann N Y Acad Sci 1137:7–17

Higgins JP et al (2003) Measuring inconsistency in meta-analyses. BMJ 327(7414):557–560

Huang Z et al (2003) Near-infrared Raman spectroscopy for optical diagnosis of lung cancer. Int J Cancer 107(6):1047–1052

Huang SJ et al (2018) Impact of age on survival of locoregional nasopharyngeal carcinoma: an analysis of the surveillance, epidemiology, and end results program database, 2004–2013. Clin Otolaryngol 43(5):1209–1218

Huedo-Medina TB et al (2006) Assessing heterogeneity in meta-analysis: q statistic or I2 index? Psychol Methods 11(2):193–206

Lean IJ et al (2009) Invited review: use of meta-analysis in animal health and reproduction: methods and applications. J Dairy Sci 92(8):3545–3565

Leslie DG et al (2012) Identification of pediatric brain neoplasms using Raman spectroscopy. Pediatr Neurosurg 48(2):109–117

Li Y (2012) Raman spectral detection and analysis of nasopharyngeal carcinoma tissue and cell lines. Fujian Normal University, Fuzhou

Li C (2017) Discrimination of nasopharyngeal carcinoma through tissue DNA label-free surface enhanced Raman spectroscopy detection. Fujian Medical University, Fuzhou

Li SX et al (2012) Detection of nasopharyngeal cancer using confocal Raman spectroscopy and genetic algorithm technique. J Biomed Opt 17(12):125003

Li Y et al (2013) Micro-Raman spectroscopy study of cancerous and normal nasopharyngeal tissues. J Biomed Opt 18(2):27003

Li Y et al (2015) Rapid detection of nasopharyngeal cancer using Raman spectroscopy and multivariate statistical analysis. Mol Clin Oncol 3(2):375–380

Lin W et al (2005) Research development and application of raman scattering technology. Chin J Light Scatt 17(2):180–186

Lin D et al (2014) Label-free blood plasma test based on surface-enhanced Raman scattering for tumor stages detection in nasopharyngeal cancer. Sci Rep 4:4751

Lin K et al (2017a) Real-time In vivo diagnosis of nasopharyngeal carcinoma using rapid fiber-optic Raman spectroscopy. Theranostics 7(14):3517–3526

Lin X et al (2017b) Noninvasive detection of nasopharyngeal carcinoma based on saliva proteins using surface-enhanced Raman spectroscopy. J Biomed Opt 22(10):1–6

Lin D et al (2018a) Autofluorescence and white light imaging-guided endoscopic Raman and diffuse reflectance spectroscopy

for in vivo nasopharyngeal cancer detection. J Biophotonics 11(4):e201700251
Lin X, Lin D, Qiu S (2018b) Detection of nasopharyngeal carcinoma based on human saliva Surface-enhanced raman spectroscopy. Spectrosc Spectr Anal 38(08):2430–2434
Lu Y (2014) Preparation of SRES substrate using cellulose as a template and its applications in the diagnosis of nasopharyngeal cancer. Fujian Agriculture and Forestry University, Fuzhou
Ma Y et al (2004) Recent developments in the determination of urinary cancer biomarkers by capillary electrophoresis. Electrophoresis 25(10–11):1473–1484
Melsen WG et al (2014) The effects of clinical and statistical heterogeneity on the predictive values of results from meta-analyses. Clin Microbiol Infect 20(2):123–129
Metz CE (1978) Basic principles of ROC analysis. Semin Nucl Med 8(4):283–298
Ming LC et al (2017) Real time near-infrared Raman spectroscopy for the diagnosis of nasopharyngeal cancer. Oncotarget 8(30):49443–49450
Molckovsky A et al (2003) Diagnostic potential of near-infrared Raman spectroscopy in the colon: differentiating adenomatous from hyperplastic polyps. Gastrointest Endosc 57(3):396–402
Ouyang H et al (2015) Rapid discrimination of malignant lesions from normal gastric tissues utilizing Raman spectroscopy system: a meta-analysis. J Cancer Res Clin Oncol 141(10):1835–1844
Pan J, Chen Q, Huang W (2011) The application value of Raman spectra for serum in early nasopharyngeal carcinoma diagnosis. China Oncol 21(09):708–712
Qiu S (2016) Non-invasive diagnosis of nasopharyngeal carcinoma based on the Raman spectroscopy. Fujian Medical University, Fuzhou
Qiu S et al (2016a) Early discrimination of nasopharyngeal carcinoma based on tissue deoxyribose nucleic acid surface-enhanced Raman spectroscopy analysis. J Biomed Opt 21(12):125003
Qiu S et al (2016b) Non-invasive detection of nasopharyngeal carcinoma using saliva surface-enhanced Raman spectroscopy. Oncol Lett 11(1):884–890
Sun L et al (2016) Preliminary study of differentiating smears from cancerous and non-cancerous nasopharyngeal tissue using confocal Raman spectroscopy. J Cancer Res Clin Oncol 142(4):823–831
Torre LA et al (2015) Global cancer statistics, 2012. CA Cancer J Clin 65(2):87–108
Wei WI, Sham JST (2005) Nasopharyngeal carcinoma. Lancet 365(9476):2041–2054
Wei KR et al (2017) Nasopharyngeal carcinoma incidence and mortality in China, 2013. Chin J Cancer 36(1):90
Whiting P et al (2003) The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 3:25
Wu S (2015) Study on smears of nasopharyngeal carcinoma tissue raman spectroscopy. Fujian Normal University, Fuzhou
Wu Q et al (2018) Assessment of the radiotherapy effect for nasopharyngeal cancer using plasma surface-enhanced Raman spectroscopy technology. Biomed Opt Express 9(7):3413–3423
Zhang J et al (2018) Accuracy of Raman spectroscopy for differentiating skin cancer from normal tissue. Medicine (Baltimore) 97(34):e12022
Zhao Q (2018) Progress of Raman spectroscopy in cancer diagnosis of human organization. Pract Oncol J 32(1):92–96
Zhao J et al (2015) Real-time Raman spectroscopy for automatic in vivo skin cancer detection: an independent validation. Anal Bioanal Chem 407(27):8373–8379