

# Triage-Tests im Rahmen der Dysphagiediagnostik

Evidenzbericht zur S3-Leitlinie Oropharyngeale  
Dysphagie infolge einer Kopf-Hals-Tumor-Erkrankung –  
Diagnostik und Therapie

A horizontal bar composed of 18 rectangular segments of varying shades of blue and grey. The word 'EVIDENZBERICHT' is written in white capital letters on a dark blue segment that spans across the middle of the bar.

**EVIDENZBERICHT**

Projekt: V21-10A

Version: 1.0

Stand: 11.10.2023

IQWiG-Berichte – Nr. 1656

DOI: 10.60584/V21-10A

# Impressum

## Herausgeber

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

## Thema

Triage-Tests im Rahmen der Dysphagiediagnostik – Evidenzbericht zur S3-Leitlinie  
Oropharyngeale Dysphagie infolge einer Kopf-Hals-Tumor-Erkrankung – Diagnostik und  
Therapie

## Auftraggeber

Bundesministerium für Gesundheit

## Datum des Auftrags

01.10.2021

## Interne Projektnummer

V21-10A

## DOI-URL

<https://dx.doi.org/10.60584/V21-10A>

## Anschrift des Herausgebers

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen  
Im Mediapark 8  
50670 Köln

Tel.: +49 221 35685-0

Fax: +49 221 35685-1

E-Mail: [berichte@iqwig.de](mailto:berichte@iqwig.de)

Internet: [www.iqwig.de](http://www.iqwig.de)

**ISSN: 1864-2500**

### **Zitiervorschlag**

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Triage-Tests im Rahmen der Dysphagiediagnostik; Evidenzbericht zur S3-Leitlinie Oropharyngeale Dysphagie infolge einer Kopf-Hals-Tumor-Erkrankung – Diagnostik und Therapie [online]. 2023 [Zugriff: TT.MM.JJJJ]. URL: <https://dx.doi.org/10.60584/V21-10A>.

### **Schlagwörter**

Diagnostik, Schluckstörungen, Kopf- und Halstumoren, Systematischer Review

### **Keywords**

Diagnosis, Deglutition Disorders, Head and Neck Neoplasms, Systematic Review

Dieser Bericht wurde ohne die Beteiligung externer Sachverständiger erstellt.

**Mitarbeiterinnen und Mitarbeiter des IQWiG**

- Alexandra Korzeczek-Opitz
- Corinna Kiefer
- Claudia Mischke
- Ulrike Paschen
- Dorothea Sow
- Andrea Tasar

# Inhaltsverzeichnis

	<b>Seite</b>
<b>Tabellenverzeichnis .....</b>	<b>vi</b>
<b>Abbildungsverzeichnis .....</b>	<b>vii</b>
<b>Abkürzungsverzeichnis.....</b>	<b>viii</b>
<b>1 Hintergrund.....</b>	<b>1</b>
<b>2 Fragestellung.....</b>	<b>2</b>
<b>3 Projektverlauf.....</b>	<b>3</b>
<b>3.1 Zeitlicher Verlauf des Projekts .....</b>	<b>3</b>
<b>3.2 Spezifizierungen und Änderungen im Projektverlauf.....</b>	<b>3</b>
<b>4 Methodik gemäß Projektskizze .....</b>	<b>5</b>
<b>4.1 Tabellarische Darstellung der Kriterien für den Studieneinschluss .....</b>	<b>5</b>
4.1.1 Kriterien für den Einschluss von Studien zur diagnostisch-therapeutischen-Kette .....	5
4.1.2 Kriterien für den Einschluss von Studien zur diagnostischen Güte.....	7
4.1.3 Einschluss von Studien, die die vorgenannten Kriterien nicht vollständig erfüllen .....	7
<b>4.2 Informationsbeschaffung.....</b>	<b>8</b>
4.2.1 Fokussierte Informationsbeschaffung von systematischen Übersichten .....	8
4.2.2 Fokussierte Informationsbeschaffung von Studien .....	8
4.2.3 Anwendung von Limitierungen auf Datenbankebene .....	9
4.2.4 Selektion relevanter Studien .....	10
<b>4.3 Informationsdarstellung und Synthese.....</b>	<b>10</b>
4.3.1 Informationsdarstellung und Synthese von Studien zur diagnostisch-therapeutischen-Kette .....	10
4.3.1.1 Darstellung der Einzelstudien .....	10
4.3.1.2 Bewertung der Studienlimitationen (Verzerrungspotenzial) .....	12
4.3.1.3 Metaanalysen .....	12
4.3.1.4 Bewertung der Qualität der Evidenz .....	13
4.3.1.4.1 Abwertung der Qualität der Evidenz.....	14
4.3.1.4.2 Aufwertung der Qualität der Evidenz .....	15
4.3.2 Informationsdarstellung und Synthese von Studien zur diagnostischen Güte....	16
4.3.2.1 Darstellung der Einzelstudien .....	16
4.3.2.2 Bewertung der Studienlimitationen (Verzerrungspotenzial) .....	17

4.3.2.3	Metaanalysen .....	18
4.3.2.4	Bewertung der Qualität der Evidenz .....	18
4.3.2.4.1	Abwertung der Qualität der Evidenz.....	19
4.3.2.4.2	Aufwertung der Qualität der Evidenz .....	20
<b>5</b>	<b>Ergebnisse .....</b>	<b>21</b>
<b>5.1</b>	<b>Ergebnisse der Informationsbeschaffung .....</b>	<b>21</b>
<b>5.2</b>	<b>Charakteristika der für die Evidenzdarstellung berücksichtigten Studien .....</b>	<b>23</b>
<b>5.3</b>	<b>Ergebnisse zur diagnostischen Güte .....</b>	<b>34</b>
5.3.1	Ergebnisse zur diagnostischen Güte: Referenzkriterium Dysphagie .....	34
5.3.1.1	Ergebnisse zum Wasserschlucktest im Vergleich zu instrumenteller Diagnostik .....	34
5.3.1.2	Ergebnisse zur klinischen Schluckuntersuchung im Vergleich zu instrumenteller Diagnostik .....	37
5.3.2	Ergebnisse zur diagnostischen Güte: Referenzkriterium Aspiration .....	40
5.3.2.1	Ergebnisse zum Wasserschlucktest im Vergleich zu instrumenteller Diagnostik .....	40
5.3.2.2	Ergebnisse zur klinischen Schluckuntersuchung im Vergleich zu instrumenteller Diagnostik .....	43
5.3.3	Ergebnisse zur diagnostischen Güte: Referenzkriterium Oralisierungseinschränkung .....	45
5.3.3.1	Ergebnisse zum Wasserschlucktest im Vergleich zu instrumenteller Diagnostik .....	45
5.3.3.2	Ergebnisse zur klinischen Schluckuntersuchung im Vergleich zu instrumenteller Diagnostik .....	47
5.3.4	Ergebnisse zum negativ prädiktiven Wert .....	49
<b>6</b>	<b>Zusammenfassung und Anmerkungen zu den Ergebnissen .....</b>	<b>52</b>
<b>6.1</b>	<b>Zusammenfassung der Ergebnisse .....</b>	<b>52</b>
<b>6.2</b>	<b>Anmerkungen zu den Ergebnissen .....</b>	<b>53</b>
<b>7</b>	<b>Literatur .....</b>	<b>55</b>
<b>8</b>	<b>Studienlisten .....</b>	<b>60</b>
<b>8.1</b>	<b>Liste der identifizierten systematischen Übersichten.....</b>	<b>60</b>
<b>8.2</b>	<b>Liste der ausgeschlossenen Publikationen mit Ausschlussgründen .....</b>	<b>60</b>
<b>Anhang A</b>	<b>Kriterien des Verzerrungspotenzials.....</b>	<b>63</b>
<b>Anhang B</b>	<b>Autorenanfragen .....</b>	<b>65</b>
<b>Anhang C</b>	<b>Suchstrategien .....</b>	<b>66</b>
<b>C.1</b>	<b>Bibliografische Datenbanken .....</b>	<b>66</b>
<b>C.2</b>	<b>Studienregister.....</b>	<b>69</b>

## Tabellenverzeichnis

	<b>Seite</b>
Tabelle 1: Übersicht über die Kriterien für den Einschluss von Studien zur diagnostisch-therapeutischen-Kette.....	6
Tabelle 2: Übersicht über die Kriterien für den Einschluss von Studien zur diagnostischen Güte .....	7
Tabelle 3: Studienpool der eingeschlossenen Studien zur diagnostischen Güte .....	23
Tabelle 4: Charakteristika der für die Evidenzdarstellung berücksichtigten Studien zur diagnostischen Güte .....	24
Tabelle 5 Charakterisierung der Index- und Referenztests in den berücksichtigten Studien zur diagnostischen Güte .....	25
Tabelle 6: Ein- / Ausschlusskriterien für Patientinnen und Patienten in den berücksichtigten Studien zur diagnostischen Güte .....	30
Tabelle 7: Charakterisierung der Studienpopulationen in den berücksichtigten Studien zur diagnostischen Güte .....	31
Tabelle 8: Evidenzprofil zur diagnostischen Güte zu Dysphagie von Wasserschlucktest im Vergleich zu instrumenteller Diagnostik .....	34
Tabelle 9: Evidenzprofil zur diagnostischen Güte zu Dysphagie von klinischer Schluckuntersuchung im Vergleich zu instrumenteller Diagnostik.....	37
Tabelle 10: Evidenzprofil zur diagnostischen Güte zu Aspiration von Wasserschlucktest im Vergleich zu instrumenteller Diagnostik .....	40
Tabelle 11: Evidenzprofil zur diagnostischen Güte zu Aspiration von klinischer Schluckuntersuchung im Vergleich zu instrumenteller Diagnostik.....	43
Tabelle 12: Evidenzprofil zur diagnostischen Güte zu Oralisierungseinschränkung von Wasserschlucktest im Vergleich zu instrumenteller Diagnostik .....	45
Tabelle 13: Evidenzprofil zur diagnostischen Güte zu Oralisierungseinschränkung von klinischer Schluckuntersuchung im Vergleich zu instrumenteller Diagnostik.....	47
Tabelle 14: Übersicht der negativ prädiktiven Werte zum Referenzkriterium Dysphagie .....	49
Tabelle 15: Übersicht der negativ prädiktiven Werte zum Referenzkriterium Aspiration .....	50
Tabelle 16: Übersicht der negativ prädiktiven Werte zum Referenzkriterium Oralisierungseinschränkung .....	51
Tabelle 17: Kriterien des Verzerrungspotenzials nach QUADAS-2 zum Referenzkriterium Dysphagie.....	63
Tabelle 18: Kriterien des Verzerrungspotenzials nach QUADAS-2 zum Referenzkriterium Aspiration.....	63
Tabelle 19: Kriterien des Verzerrungspotenzials nach QUADAS-2 zum Referenzkriterium Oralisierungseinschränkung .....	64
Tabelle 20: Übersicht über Autorenanfragen .....	65

## Abbildungsverzeichnis

	<b>Seite</b>
Abbildung 1: Ergebnis der Informationsbeschaffung .....	22

**Abkürzungsverzeichnis**

<b>Abkürzung</b>	<b>Bedeutung</b>
AWMF	Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e. V.
BMG	Bundesministerium für Gesundheit
DGPP	Deutsche Gesellschaft für Phoniatrie und Pädaudiologie e. V.
DVG	Digitale-Versorgung-Gesetz
FEES	fiberoptic endoscopic evaluation of swallowing (fiberoptisch endoskopische Evaluation des Schluckvorganges)
FraMaDySc	Frankfurt/Marburg Dysphagia Screening
GRADE	Grading of Recommendations Assessment, Development and Evaluation
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ITT	Intention to treat
KI	Konfidenzintervall
MASA-C	Mann Assessment of Swallowing Ability-Cancer
MWD	Mittelwertdifferenz
NPV	negative predictive value (negativer prädiktiver Wert)
OR	Odds Ratio
PICO	Population, Intervention, Comparison, Outcome
QUADAS	Quality Assessment of Diagnostic Accuracy Studies
RCT	Randomized controlled Trial (randomisierte kontrollierte Studie)
RD	absolute Risikodifferenz
SMD	standardisierte Mittelwertdifferenz
ROC	Receiver Operating Characteristic
SÜ	systematische Übersicht
VFSS	videofluoroscopic swallowing study (videofluoroskopische Schluckuntersuchung)
WST	Wasserschlucktest

## **1 Hintergrund**

Auf Basis des Digitale-Versorgung-Gesetzes (DVG) kann die Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) dem Bundesministerium für Gesundheit (BMG) Leitlinien vorschlagen, bei denen das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) gemäß SGB V (§§ 139a Abs. 3 Nr. 3, 139b Abs. 6) deren Entwicklung oder Aktualisierung mit Evidenzrecherchen unterstützen soll [1].

Der vorliegende Auftrag umfasst die Beantwortung von Fragestellungen zur Entwicklung der interdisziplinären S3-Leitlinie „Oropharyngeale Dysphagie infolge einer Kopf-Hals-Tumor-Erkrankung – Diagnostik und Therapie“ (Registernummer der AWMF: 049-016) [2].

## **2 Fragestellung**

Ziel des Evidenzberichts ist die Darstellung der Evidenz bezüglich der Effekte eines Triage-Tests (Wasserschlucktest, klinische Schluckuntersuchung oder Fragebogen zur subjektiven Einschätzung des Schluckvermögens) vor Einsatz instrumenteller Diagnostik bei Testpositiven im Vergleich zum alleinigen Einsatz instrumenteller Diagnostik zur Indikation für eine Dysphagietherapie bei Patientinnen und Patienten mit Kopf-Hals-Tumor während / nach Strahlen-(Chemo-)therapie oder nach Tumorresektion.

Als Baustein der Fragestellung wird zudem die Evidenz bezüglich der diagnostischen Güte von Wasserschlucktest, klinischer Schluckuntersuchung oder Fragebogen zur subjektiven Einschätzung des Schluckvermögens im Vergleich zur instrumentellen Diagnostik bei Patientinnen und Patienten mit Kopf-Hals-Tumor während / nach Strahlen-(Chemo-)therapie oder nach Tumorresektion dargestellt.

### **3 Projektverlauf**

#### **3.1 Zeitlicher Verlauf des Projekts**

Das IQWiG wurde am 01.10.2021 vom BMG beauftragt, die Deutsche Gesellschaft für Phoniatrie und Pädaudiologie e. V. (DGPP) bei der Entwicklung einer interdisziplinären S3-Leitlinie „Oropharyngeale Dysphagie infolge einer Kopf-Hals-Tumor-Erkrankung – Diagnostik und Therapie“ [2] zu unterstützen.

Insgesamt wurden 5 Population-Intervention-Comparison-Outcome(PICO)-Fragen von den Leitlinienkoordinatorinnen in Abstimmung mit Patientenvertreterinnen und -vertretern und mit Beratung durch die AWMF formuliert, zu denen das IQWiG jeweils auf Basis einer internen Projektskizze einen Evidenzbericht erstellte. Die PICO-Fragen wurden im Vorfeld mit den Leitlinienkoordinatorinnen, den Ansprechpartnerinnen der AWMF und Ansprechpartnerinnen und -partnern des IQWiG konkretisiert. Die Evidenzberichte wurden an die Koordinatorinnen der Leitlinienerstellung, an die zuständige Ansprechpartnerin für die Leitlinie bei der AWMF sowie an das BMG übermittelt.

Nach Abschluss aller Evidenzberichte für diesen Auftrag wurden diese zusammen an das BMG übermittelt und 4 Wochen später auf der Website des IQWiG veröffentlicht.

#### **3.2 Spezifizierungen und Änderungen im Projektverlauf**

Es ergaben sich folgende Spezifizierungen im Evidenzbericht:

- Abschnitt 4.1.2: Es wurden alle Querschnitts- und Kohortenstudien eingeschlossen, unabhängig davon, ob sie prospektiv geplant oder retrospektiv durchgeführt wurden.
- Abschnitt 4.3.2.1: Die Ergebnisse zu den in den Studien berichteten Vierfeldertafel-Daten wurden im Bericht in Forest Plots dargestellt, sofern mindestens 2 Studien vorlagen. Neben der Darstellung von Sensitivität und Spezifität, welche die Validität eines Tests betreffen, ist der negative prädiktive Wert (NPV) für die ärztliche Entscheidung zum Einsatz eines Triage-Tests von Bedeutung. Der NPV gibt den Anteil der Personen mit richtig negativem Testergebnis, d. h. die Personen, die tatsächlich frei von der gesuchten Krankheit sind, im Verhältnis zu allen Personen mit negativem Testergebnis an. Er wird durch die Angaben zur Sensitivität und Spezifität des Tests sowie durch die Prävalenz der Krankheit in der getesteten Grundgesamtheit (Vortestwahrscheinlichkeit) definiert (Bayes'sches Theorem).
- Abschnitt 4.3.2.4.1: Der Wertebereich der Sensitivität bzw. Spezifität wurde mittels einer Grenze bei 50 % und einer bei 80 % für Spezifität bzw. 90 % für Sensitivität in 3 Sektoren aufgeteilt: 0 % bis 50 %, 50 % bis 80 % bzw. 90% und 80 % bzw. 90% bis 100 %. Wegen fehlender Genauigkeit wurde um 1 Stufe abgewertet, wenn das 95 %-Konfidenzintervall

(KI) der (gepoolten) Schätzung für die Sensitivität bzw. die Spezifität in 2 aneinander angrenzenden Sektoren lag. Wegen fehlender Genauigkeit wurde um 2 Stufen abgewertet, wenn das 95 %-KI der (gepoolten) Schätzung für die Sensitivität bzw. die Spezifität sich über alle 3 Sektoren erstreckte. Eine Sensitivität höher als 90 % bzw. eine Spezifität höher als 80 % wurde als empfehlenswert für die klinische Praxis angenommen (im Weiteren als Empfehlungsgrenze bezeichnet). Es wurde mit 90 % eine sehr hohe Grenze für die Sensitivität festgelegt, da besonders die Folgen einer falsch-negativen Diagnose anhand des Triage-Tests schwerwiegend für die Patientinnen und Patienten sein können, wenn eine behandlungsbedürftige Dysphagie möglicherweise unerkannt bleibt. Eine Sensitivität bzw. eine Spezifität niedriger als 50 % wird als unbrauchbar für die klinische Praxis betrachtet, da in diesem Fall das Testergebnis ein Zufallsergebnis sein könnte.

## **4 Methodik gemäß Projektskizze**

Die folgenden Abschnitte geben den Wortlaut der Berichtsmethodik aus der Projektskizze wieder. Über diese Methodik hinausgehende Spezifizierungen oder Änderungen der Methoden im Projektverlauf werden in Abschnitt 3.2 erläutert. Im folgenden Text wird an den entsprechenden Stellen auf diesen Abschnitt verwiesen.

Die PICO-Fragestellung wird durch die Darstellung von Evidenzprofilen beantwortet. Die Erstellung der Evidenzprofile erfolgt auf Grundlage der methodischen Vorgaben von Grading of Recommendations Assessment, Development and Evaluation (GRADE) [3] und ist konform zu den Allgemeinen Methoden [4]. Das PICO-Schema wurde von der Leitliniengruppe zur interdisziplinären S3-Leitlinie „Oropharyngeale Dysphagie infolge einer Kopf-Hals-Tumor-Erkrankung – Diagnostik und Therapie“ (Registernummer der AWMF: 049-016) [2] festgelegt.

Die Fragestellung dieses Berichts lässt sich anhand von vergleichenden Interventionsstudien der gesamten diagnostisch-therapeutischen-Kette bewerten. Die Einschlusskriterien zu diesem Vorgehen finden sich in Abschnitt 4.1.1.

Zudem wird der erste Baustein der Fragestellung, die diagnostische Güte, untersucht. Unter Annahme eines Nutzens der Therapie auf Basis der Referenzdiagnostik bei Dysphagie [5], ist es auch auf diesem Weg möglich, einen Vor- oder Nachteil des Einsatzes eines Triage-Tests zu zeigen. Die Einschlusskriterien zur Untersuchung der diagnostischen Güte finden sich in Abschnitt 4.1.2.

### **4.1 Tabellarische Darstellung der Kriterien für den Studieneinschluss**

#### **4.1.1 Kriterien für den Einschluss von Studien zur diagnostisch-therapeutischen-Kette**

In der folgenden Tabelle 1 sind die Kriterien aufgelistet, die Studien erfüllen müssen, um in die Evidenzdarstellung eingeschlossen zu werden.

Tabelle 1: Übersicht über die Kriterien für den Einschluss von Studien zur diagnostisch-therapeutischen-Kette

Einschlusskriterien	
E1	Population: Patientinnen und Patienten ( $\geq 18$ Jahre) mit Kopf-Hals-Tumor (Plattenepithelkarzinom) während / nach Strahlen-(Chemo-)Therapie oder nach Tumorresektion
E2	Prüfintervention: Triage-Test (Wasserschlucktest, klinische Schluckuntersuchung oder Fragebogen zur subjektiven Einschätzung des Schluckvermögens <sup>a</sup> ) vor Einsatz einer instrumentellen Diagnostik (FEES oder VFSS) bei Testpositiven
E3	Vergleichsintervention: instrumentelle Diagnostik (FEES oder VFSS)
E4	kritische Endpunkte: Aspirationspneumonie, Aspirationen (graduiert nach PAS), Oralisierungseinschränkungen (z. B. FOIS, Kostformanpassung, Sondenversorgung), Schluckstörung, Malnutrition (z. B. BMI, ungewollter Gewichtsverlust, reduzierte Muskelmasse)
E5	Studientypen: RCTs und ggf. nicht randomisierte vergleichende Studien (schrittweiser Einschluss von niedrigeren Evidenzstufen bei ungenügender Zahl und / oder Qualität: quasirandomisierte kontrollierte Studien, prospektive vergleichende Kohortenstudien, retrospektive vergleichende Kohortenstudien mit zeitlich paralleler Kontrollgruppe, retrospektive vergleichende Kohortenstudien mit nicht zeitlich paralleler Kontrollgruppe)
E6	Publikationssprache: Deutsch oder Englisch
E7	Vollpublikation verfügbar <sup>b</sup>
E8	Publikationszeitraum ab 2011
<p>a. Es werden folgende Fragebogen untersucht: EAT-10, DHI, SSQ, MDADI, SWALQoL, EORTC QLQ-C30.  b. Als Vollpublikation gilt in diesem Zusammenhang auch ein Studienbericht gemäß ICH E3 [6] oder ein Bericht über die Studie, der den Kriterien des CONSORT- [7], TREND- [8] oder STROBE-Statements [9] genügt und eine Bewertung der Studie ermöglicht, sofern die in diesen Dokumenten enthaltenen Informationen zur Studienmethodik und zu den Studienergebnissen nicht vertraulich sind.</p> <p>BMI: Body-Mass-Index; CONSORT: Consolidated Standards of Reporting Trials; DHI: Dysphagia-Handycap-Index; EAT-10: Eating Assessment Tool; EORTC QLQ-C30: European Organization for Research and Treatment of Cancer Core Quality of Life Questionnaire - Core Questionnaire; FEES: fiberoptisch endoskopische Evaluation des Schluckvorganges; FOIS: Functional Oral Intake Scale; ICH: International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use; MDADI: M. D. Anderson Dysphagia Inventory; PAS: Penetrations-Aspirations Skala; RCT: randomisierte kontrollierte Studie; SSQ: Sydney Swallowing Questionnaire; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology; SWALQoL: quality of life in swallowing disorders; TREND: Transparent Reporting of Evaluations with Nonrandomized Designs; VFSS: videofluoroskopische Schluckuntersuchung</p>	

#### 4.1.2 Kriterien für den Einschluss von Studien zur diagnostischen Güte

In der folgenden Tabelle 2 sind die Kriterien aufgelistet, die Studien erfüllen müssen, um in die Evidenzdarstellung eingeschlossen zu werden.

Tabelle 2: Übersicht über die Kriterien für den Einschluss von Studien zur diagnostischen Güte

Einschlusskriterien	
DN1	Population: Patientinnen und Patienten (≥ 18 Jahre) mit Kopf-Hals-Tumor (Plattenepithelkarzinom) während / nach Strahlen-(Chemo-)Therapie oder nach Tumorresektion
DN2	Indextests: Wasserschlucktest, klinische Schluckuntersuchung oder Fragebogen zur subjektiven Einschätzung des Schluckvermögens <sup>a</sup>
DN3	Referenzstandard: instrumentelle Diagnostik (FEES oder VFSS)
DN4	Zielgrößen: personenbezogene Vierfeldertafel-Daten zur Berechnung der diagnostischen Güte (z. B. Sensitivität, Spezifität)
DN5	Studientyp: prospektiv geplante Querschnitts- und Kohortenstudien <sup>b</sup>
DN6	Publikationssprache: Deutsch oder Englisch
DN7	Vollpublikation verfügbar <sup>c</sup>
DN8	Publikationszeitraum ab 2011
<p>a. Es werden folgende Fragebogen untersucht: EAT-10, DHI, SSQ, MDADI, SWALQoL, EORTC QLQ-C30                      b. Zentrale Aspekte der Ergebnissicherheit bei Studien zur diagnostischen Güte sind der konsekutive Einschluss der Patientinnen und Patienten und die Dokumentation fehlender Werte. Bei ausreichender Anzahl und Qualität werden nur die Studien eingeschlossen, die diese Voraussetzungen erfüllen.                      c. Als Vollpublikation gilt in diesem Zusammenhang auch ein Bericht über die Studie, der den Kriterien des STARD- [10] oder STROBE-Statements [9] genügt und eine Bewertung der Studie ermöglicht, sofern die in diesen Dokumenten enthaltenen Informationen zur Studienmethodik und zu den Studienergebnissen nicht vertraulich sind.</p> <p>DHI: Dysphagia-Handycap-Index; EAT-10: Eating Assessment Tool; EORTC QLQ-C30: European Organization for Research and Treatment of Cancer Core Quality of Life Questionnaire - Core Questionnaire; FEES: fiberoptisch endoskopische Evaluation des Schluckvorganges; MDADI: M. D. Anderson Dysphagia Inventory; STARD: Standards for the Reporting of Diagnostic Accuracy; SSQ: Sydney Swallowing Questionnaire; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology; SWALQoL: quality of life in swallowing disorders; VFSS: videofluoroskopische Schluckuntersuchung</p>	

(Zu diesem Vorgehen gab es eine Spezifizierung im Projektverlauf, siehe Abschnitt 3.2.)

#### 4.1.3 Einschluss von Studien, die die vorgenannten Kriterien nicht vollständig erfüllen

Für die Einschlusskriterien E1 bzw. DN1 (Population), E2 (Prüfintervention, bezogen auf die Interventionsgruppe der Studie) bzw. DN2 (Indextests) und E3 (Vergleichsintervention, bezogen auf die Vergleichsgruppe der Studie) bzw. DN3 (Referenztests) reicht es aus, wenn bei mindestens ca. 80 % der eingeschlossenen Patientinnen und Patienten diese Kriterien erfüllt sind. Liegen für solche Studien Subgruppenanalysen für Patientinnen und Patienten vor, die die Einschlusskriterien erfüllen, wird auf diese Analysen zurückgegriffen. Studien, bei

denen die Einschlusskriterien E1, E2 und E3 bzw. DN1, DN2 und DN3 bei weniger als ca. 80 % erfüllt sind, können ggf. unberücksichtigt bleiben.

## **4.2 Informationsbeschaffung**

### **4.2.1 Fokussierte Informationsbeschaffung von systematischen Übersichten**

Parallel zur Erstellung der Projektskizze erfolgte eine Recherche nach systematischen Übersichten (SÜs) in MEDLINE (umfasst auch die Cochrane Database of Systematic Reviews), der International Health Technology Assessment HTA Database sowie auf den Websites des National Institute for Health and Care Excellence (NICE) und der Agency for Healthcare Research and Quality (AHRQ).

Die Suche fand am 03.04.2023 statt. Die Suchstrategien für die Suche in bibliografischen Datenbanken finden sich in Anhang C.

Die Selektion erfolgte durch 1 Person und wurde anschließend von einer 2. Person überprüft. Diskrepanzen wurden durch Diskussion zwischen beiden aufgelöst.

Es wird geprüft, ob mindestens eine hochwertige und aktuelle SÜ infrage kommt, deren Informationsbeschaffung als Grundlage für die Evidenzdarstellung verwendet werden kann (im Folgenden: Basis-SÜ). Dafür erfolgt eine Bewertung der Qualität der Informationsbeschaffung dieser SÜ(s). Kann mindestens eine diesbezüglich hochwertige und aktuelle Basis-SÜ identifiziert werden, werden die zugrunde liegenden Studien bzw. Dokumente von 1 Person auf deren Relevanz für die vorliegende Evidenzdarstellung geprüft und das Ergebnis von einer 2. Person überprüft. Bewertungen der eingeschlossenen Studien oder die Datenextraktion werden nicht übernommen.

Die finale Entscheidung, ob und wenn ja welche SÜ(s) als Basis-SÜ(s) für die Evidenzdarstellung herangezogen werden, erfolgt nach Erstellung der Projektskizze anhand der darin festgelegten Kriterien. In jedem Fall werden die Referenzlisten der identifizierten SÜ(s) hinsichtlich relevanter Primärstudien gesichtet (siehe Abschnitt 4.2.2).

### **4.2.2 Fokussierte Informationsbeschaffung von Studien**

Für den Fall, dass mindestens 1 SÜ als Basis-SÜ für die Informationsbeschaffung des Evidenzberichts verwendet werden kann (siehe Abschnitt 4.2.1), wird diese für die Informationsbeschaffung von Studien für den von der Übersicht abgedeckten Zeitraum herangezogen. Dieser Teil der Informationsbeschaffung wird ergänzt um eine systematische Recherche nach relevanten Studien bzw. Dokumenten für den nicht von der Übersicht abgedeckten Zeitraum. Gegebenenfalls wird auf eine ergänzende fokussierte Informationsbeschaffung von Studien ganz verzichtet.

Für den Fall, dass keine Basis-SÜ identifiziert werden kann, findet eine systematische Recherche für den gesamten relevanten Zeitraum statt.

Folgende primäre und weitere Informationsquellen sowie Suchtechniken werden dabei berücksichtigt:

### **Primäre Informationsquellen**

- bibliografische Datenbanken
  - MEDLINE
  - Cochrane Central Register of Controlled Trials
- Studienregister (es erfolgt eine Einschränkung auf Studienergebnisse)
  - U.S. National Institutes of Health. ClinicalTrials.gov
  - World Health Organization. International Clinical Trials Registry Platform Search Portal

### **Weitere Informationsquellen und Suchtechniken**

- Anwendung weiterer Suchtechniken
  - Sichten von Referenzlisten identifizierter systematischer Übersichten
  - Identifizieren von Studienregistereinträgen zu eingeschlossenen Studien
- Autorenanfrage(n)

### **4.2.3 Anwendung von Limitierungen auf Datenbankebene**

#### **Fokussierte Informationsbeschaffung von systematischen Übersichten**

Die Suchen wurden auf das Publikationsdatum ab Januar 2018 eingeschränkt. Die MEDLINE-Suchstrategie enthält Limitierungen auf deutsch- und englischsprachige Publikationen [4] sowie auf Humanstudien.

#### **Fokussierte Informationsbeschaffung von Studien**

Die Suchen werden auf das Publikationsdatum ab Januar 2011 eingeschränkt. Mit der MEDLINE-Suchstrategie werden folgende Publikationstypen ausgeschlossen: Kommentare und Editorials, da diese in der Regel keine Studien enthalten [11]. Außerdem enthalten die Suchstrategien Limitierungen auf deutsch- und englischsprachige Publikationen [4] sowie auf Humanstudien (MEDLINE). In der Cochrane Central Register of Controlled Trials Suche werden Einträge aus Studienregistern ausgeschlossen.

#### **4.2.4 Selektion relevanter Studien**

##### **Selektion relevanter Studien bzw. Dokumente aus den Ergebnissen der bibliografischen Recherche**

Duplikate werden mit Hilfe des Literaturverwaltungsprogrammes EndNote entfernt. Die in bibliografischen Datenbanken identifizierten Treffer werden in einem 1. Schritt anhand ihres Titels und, sofern vorhanden, Abstracts in Bezug auf ihre potenzielle Relevanz bezüglich der Einschlusskriterien (siehe Tabelle 1 und Tabelle 2) bewertet. Als potenziell relevant erachtete Dokumente werden in einem 2. Schritt anhand ihres Volltextes auf Relevanz geprüft. Beide Schritte erfolgen durch 2 Personen unabhängig voneinander. Diskrepanzen werden durch Diskussion zwischen beiden aufgelöst.

##### **Selektion relevanter Studien bzw. Dokumente aus weiteren Informationsquellen**

Rechercheergebnisse aus den folgenden Informationsquellen werden von 1 Person auf Studien gesichtet:

- Studienregister
- Referenzlisten identifizierter systematischer Übersichten

Die identifizierten Studien werden dann auf ihre Relevanz geprüft. Der gesamte Prozess wird anschließend von einer 2. Person überprüft. Diskrepanzen werden durch Diskussion zwischen beiden aufgelöst.

#### **4.3 Informationsdarstellung und Synthese**

##### **4.3.1 Informationsdarstellung und Synthese von Studien zur diagnostisch-therapeutischen-Kette**

###### **4.3.1.1 Darstellung der Einzelstudien**

Alle für den Evidenzbericht notwendigen Informationen werden aus den Unterlagen zu den eingeschlossenen Studien in standardisierte Tabellen extrahiert. Ergeben sich im Abgleich der Informationen aus unterschiedlichen Dokumenten zu einer Studie (aber auch aus multiplen Angaben zu einem Aspekt innerhalb eines Dokumentes selbst) Diskrepanzen, die auf die Interpretation der Ergebnisse erheblichen Einfluss haben könnten, wird dies an den entsprechenden Stellen des Berichts dargestellt.

Die Extraktion erfolgt durch eine Person und wird von einer 2. Person auf Grundlage der Studien kontrolliert.

Die Ergebnisse zu den in den Studien berichteten kritischen Endpunkten werden im Bericht vergleichend beschrieben.

Das Vorgehen zu der Bewertung des Einflusses der Studienlimitationen auf die berichtsrelevanten Ergebnisse wird in Abschnitt 4.3.1.2 endpunktspezifisch pro Studie beschrieben. Nach dieser Bewertung werden die Informationen zu den ausgewählten Endpunkten in Evidenzprofilen zusammengeführt und analysiert [12]. Wenn möglich werden über die Gegenüberstellung der Ergebnisse der Einzelstudien hinaus die in Abschnitt 4.3.1.3 beschriebenen Verfahren eingesetzt.

Für binäre Daten wird als relative Effektschätzung primär das Odds Ratio (OR) herangezogen. Dabei wird im Fall von 0 Ereignissen in einem Studienarm bei der Berechnung von Effekt und KI der Korrekturfaktor 0,5 in beiden Studienarmen verwendet. Für absolute Effekte wird vorrangig die absolute Risikodifferenz (RD) verwendet. Das Basisrisiko wird in der Regel durch den Median des Risikos der Kontrollgruppe in den eingeschlossenen Einzelstudien bestimmt. Auf Grundlage dieses angenommenen Basisrisikos werden mithilfe der relativen Gesamtschätzung der Metaanalyse das absolute Risiko in der Interventionsgruppe und die Risikodifferenz errechnet. Zur Abschätzung der quantitativen Ergebnisunsicherheit werden ergänzend die oberen und unteren KI-Grenzen der RD basierend auf den KI-Grenzen der relativen Gesamtschätzung der Metaanalyse berechnet. Wird keine Metaanalyse durchgeführt, erfolgt eine Berechnung der RD pro Studie durch die Risiken in den beiden Behandlungsgruppen; als zugehöriges KI wird standardmäßig dasjenige nach der Wilson-Score-Methode [13] angegeben. Sollte das Ergebnis des CSZ-Tests [14] qualitativ nicht zum Ergebnis des KI nach der Wilson-Score-Methode, sondern zum Wald-KI passen, wird dieses angegeben.

Für stetige Daten wird primär die Mittelwertdifferenz (MWD) herangezogen. Falls notwendig (z. B. wenn verschiedene Skalen gepoolt werden sollen oder dies zur Bewertung der Qualität der Evidenz benötigt wird), wird außerdem eine standardisierte Mittelwertdifferenz (SMD, z. B. Hedges' g) angegeben.

Ergebnisse können ggf. im Evidenzbericht unberücksichtigt bleiben, wenn ein großer Anteil der in die Auswertung eigentlich einzuschließenden Personen nicht in der Auswertung berücksichtigt worden ist. Für die Entscheidung hierüber wird sich an einem Anteil von ca. 70 % orientiert, die in der Auswertung mindestens berücksichtigt sein sollten.

Die Ergebnisse können ggf. auch dann unberücksichtigt bleiben, wenn der Unterschied der Anteile nicht berücksichtigter Personen zwischen den Gruppen sehr groß ist. Als Orientierung für die Entscheidung dient ein Unterschied von 15 Prozentpunkten.

Die festgelegten Grenzen von 70 % und 15 Prozentpunkte können in begründeten Einzelfällen projektspezifisch verändert werden.

#### 4.3.1.2 Bewertung der Studienlimitationen (Verzerrungspotenzial)

Der Einfluss möglicher Studienlimitationen auf die Ergebnisse wird zunächst endpunktspezifisch pro Studie insbesondere anhand der unten aufgeführten Kriterien bewertet. Dazu erfolgt jeweils eine Bewertung mit „adäquat“, „unklar“ oder „nicht adäquat“. Eine Bewertung mit „unklar“ erfolgt grundsätzlich dann, wenn keine bzw. keine ausreichenden Angaben zur adäquaten Bewertung zur Verfügung stehen. Im Evidenzprofil werden abschließend die Studienlimitationen studienübergreifend und endpunktspezifisch durch die Bewertungen „nicht schwerwiegend“, „schwerwiegend“, „sehr schwerwiegend“ dargestellt.

- adäquate Erzeugung der Randomisierungssequenz (bei randomisierten Studien)
- Verdeckung der Gruppenzuteilung (bei randomisierten Studien)
- zeitliche Parallelität der Gruppen (bei nicht randomisierten Studien)
- Vergleichbarkeit der Gruppen beziehungsweise adäquate Berücksichtigung prognostisch relevanter Faktoren (bei nicht randomisierten Studien)
- Verblindung
- Umsetzung des Intention-to-treat(ITT)-Prinzips
- ergebnisunabhängige Berichterstattung

#### 4.3.1.3 Metaanalysen

Die geschätzten Effekte und KIs aus den Studien werden mittels Forest Plots zusammenfassend dargestellt. Anschließend wird die Heterogenität des Studienpools anhand des statistischen Tests auf Vorliegen von Heterogenität [15] untersucht. Ergibt der Heterogenitätstest ein statistisch nicht signifikantes Ergebnis ( $p \geq 0,05$ ), wird davon ausgegangen, dass die Schätzung eines gemeinsamen (gepoolten) Effekts sinnvoll ist. Im Fall von mindestens 5 Studien erfolgt die Metaanalyse mithilfe des Modells mit zufälligen Effekten gemäß der Methode nach Knapp und Hartung unter Verwendung des Heterogenitätsschätzers nach Paule und Mandel [16]. Als Ergebnis wird der gemeinsame Effekt inklusive KI dargestellt. Weil die Heterogenität im Fall weniger Studien nicht verlässlich geschätzt werden kann, werden bei 4 oder weniger Studien gegebenenfalls Modelle mit festem Effekt verwendet. Sind in einer Metaanalyse nur 2 Studien enthalten, so wird aus pragmatischen Gründen regelhaft ein Modell mit festem Effekt gewählt, wenn keine deutlichen Gründe dagegensprechen. Bei 3 und 4 Studien erfolgt vorzugsweise eine Metaanalyse nach der Methode von Knapp und Hartung, sofern das KI informativ ist und kein Widerspruch zu einer qualitativen Zusammenfassung besteht. Andernfalls wird ein Modell mit festem Effekt verwendet, wenn keine deutlichen Gründe dagegensprechen. Ist ein Modell mit festem Effekt nicht vertretbar, bleibt es bei der Schätzung nach Knapp und Hartung.

Ergibt der Heterogenitätstest ein statistisch signifikantes Ergebnis ( $p < 0,05$ ), wird untersucht, welche Faktoren diese Heterogenität möglicherweise verursachen. Dazu zählen methodische und klinische Faktoren, sogenannte Effektmodifikatoren. Kann die Heterogenität durch diese Faktoren erklärt werden, wird das Poolen der Effekte ggf. auf Subgruppen beschränkt [17]. Kann die Heterogenität nicht erklärt werden, so wird, sofern möglich, ebenfalls ein gemeinsamer (gepoolter) Effekt berechnet. Dabei ist zu beachten, dass ein gepoolter Effekt bei bedeutsamer Heterogenität möglicherweise nicht sinnvoll interpretierbar ist. Von der Berechnung eines gepoolten Effekts wird abgesehen, falls sich die KIs der eingehenden Studien nur gering oder gar nicht überlappen und gleichzeitig die Effektschätzungen der Studien in unterschiedliche Richtungen weisen.

#### **4.3.1.4 Bewertung der Qualität der Evidenz**

Alle für den Evidenzbericht relevanten Ergebnisse werden hinsichtlich der Beeinflussung durch Studienlimitationen, des Risikos von Publikationsbias, der Genauigkeit und Konsistenz der Effekte, der Übertragbarkeit und ggf. hinsichtlich möglicher Aspekte zur Aufwertung der Qualität der Evidenz überprüft. Für jeden für die Entscheidung kritischen Endpunkt wird eine endpunktbezogene studienübergreifende Aussage zur Qualität der Evidenz in 4 Abstufungen bezüglich des jeweiligen Vertrauens in die Effektschätzung getroffen und diese in eine der Stufen hohe, moderate, niedrige und sehr niedrige Qualität der Evidenz entsprechend der GRADE-Guideline eingeteilt [18,19]:

- Das Vertrauen in die Effektschätzung ist hoch: Eine hohe Qualität der Evidenz bedeutet, dass der wahre Effekt sehr sicher nahe bei der Effektschätzung liegt.
- Das Vertrauen in die Effektschätzung ist mäßig: Eine moderate Qualität der Evidenz bedeutet, dass der wahre Effekt wahrscheinlich nahe bei der Effektschätzung liegt, aber die Möglichkeit besteht, dass er relevant verschieden ist.
- Das Vertrauen in die Effektschätzung ist begrenzt: Eine niedrige Qualität der Evidenz bedeutet, dass der wahre Effekt durchaus relevant verschieden zur Effektschätzung sein kann.
- Das Vertrauen in die Effektschätzung ist sehr gering: Eine sehr niedrige Qualität der Evidenz bedeutet, dass der wahre Effekt wahrscheinlich relevant verschieden von der Effektschätzung ist.

Die Bewertung erfolgt durch 2 Personen unabhängig voneinander. Diskrepanzen werden durch Diskussion zwischen beiden aufgelöst.

Ergebnissen aus RCTs wird im 1. Bewertungsschritt eine hohe, Ergebnissen aus Beobachtungsstudien eine niedrige Qualität der Evidenz attestiert [18]. Von dieser rein designbedingten Einschätzung ausgehend, wird bei Studienlimitationen, einem Risiko von

Publikationsbias, fehlender Genauigkeit der Effekte, inkonsistenten Effekten oder eingeschränkter Übertragbarkeit die Qualität der Evidenz abgewertet sowie bei (sehr) großen Effekten, einer Dosis-Wirkungs-Beziehung oder, wenn die Berücksichtigung aller potenziellen Confounder zu einer Effekterhöhung / Verstärkung des beobachteten Ergebnisses führen würde, aufgewertet. Unterstützend für diese Bewertungsschritte können Sensitivitätsanalysen durchgeführt werden. Alle Ab- und Aufwertungen der Qualität der Evidenz werden durch Fußnoten gekennzeichnet.

#### **4.3.1.4.1 Abwertung der Qualität der Evidenz**

##### **A: Studienlimitationen**

Die Qualität der Evidenz wird aufgrund von schwerwiegenden oder sehr schwerwiegenden Studienlimitationen um 1 oder 2 Stufen abgewertet, wenn starke Limitierungen in einem oder mehreren der in Abschnitt 4.3.1.2 genannten Kriterien das Vertrauen in die Effektschätzung beeinträchtigen. Die endpunktbezogene studienübergreifende Bewertung der Studienlimitationen erfolgt unter Berücksichtigung des Einflusses der einzelnen Studien auf die Effektschätzung für jeden Endpunkt [20].

##### **B: Inkonsistente (heterogene) Effekte**

Die Qualität der Evidenz wird bei Inkonsistenz (unerklärter Heterogenität) zwischen Studienergebnissen um 1 oder 2 Stufen abgewertet. Die Einschätzung einer möglichen Heterogenität erfolgt anhand der in Abschnitt 4.3.1.3 beschriebenen Kriterien.

Falls für binäre Daten Metaanalysen für 2 Effektmaße durchgeführt wurden und eine der beiden einen statistisch signifikanten Gruppenunterschied anzeigt, wird die Inkonsistenz anhand dieser Metaanalyse bewertet. Ist der Gruppenunterschied für beide Effektmaße nicht statistisch signifikant, wird die Bewertung in der Regel anhand des relativen Effektmaßes vorgenommen.

Bei stetigen Daten wird die Bewertung in der Regel anhand der Metaanalyse der Mittelwertdifferenz vorgenommen. Falls die fehlende Genauigkeit anhand der standardisierten Mittelwertdifferenz bewertet wurde, wird die Inkonsistenz auch anhand dieser bewertet.

Eine Abwertung um 1 Stufe erfolgt, wenn der Heterogenitätstest ein signifikantes Ergebnis liefert. Um 2 Stufen kann z. B. abgewertet werden, wenn aufgrund der Heterogenität der Studienergebnisse keine gepoolte Effektschätzung berechnet wird.

### **C: Indirektheit**

Die Qualität der Evidenz wird bei starker oder sehr starker Indirektheit um 1 oder 2 Stufen abgewertet. Indirektheit kann auf Abweichungen zwischen dem PICO und den Einschlusskriterien der Studien basieren [21].

### **D: Publikationsbias**

Die Qualität der Evidenz wird um 1 Stufe abgewertet, wenn ein Publikationsbias wahrscheinlich oder sehr wahrscheinlich ist.

Kriterien für die Annahme eines Publikationsbias umfassen eine Evidenzgrundlage, welche vorrangig auf kleinen Studien und wenigen Ereignissen beruht, welche einen beobachteten Zusammenhang zwischen Größe der Effektschätzung und Studiengröße (oder Genauigkeit der Effekte) zeigt, welche Studienabbrüche auf der Grundlage von hohen Effektschätzungen aufweist oder welche Auffälligkeiten in der gesichteten Datenlage erkennen lässt [22].

### **E: Fehlende Genauigkeit der Effekte**

Die Qualität der Evidenz wird wegen geringer oder sehr geringer Genauigkeit der Effekte um 1 oder 2 Stufen abgewertet. Maßgeblich hierfür ist neben der Lage und Breite des 95 %-KI auch die Wahl des Metaanalysemodells (zufällige Effekte oder fester Effekt) (siehe Abschnitt 4.3.1.3). Außerdem können sehr kleine Fallzahlen zu einer Abwertung wegen fehlender Genauigkeit führen [23]. Bei ausreichender Patientenzahl wird regelhaft nicht abgewertet, wenn das KI eines Effekts (relativ oder absolut im Fall binärer Daten bzw. Mittelwertdifferenz oder standardisierte Mittelwertdifferenz im Fall stetiger Daten) einen signifikanten Unterschied anzeigt.

Andernfalls erfolgt eine Abwertung wegen fehlender Präzision, falls das KI (des relativen Effekts bei binären Daten bzw. Hedges'  $g$  bei stetigen Daten) einen mittelgroßen Effekt und gleichzeitig auch den Nulleffekt überdeckt. Bei binären Daten wird unter einem mittelgroßen Effekt eine Halbierung oder eine Verdopplung (0,5 oder 2) des relativen Effekts verstanden. Bei stetigen Daten wird von einem mittelgroßen Effekt ausgegangen, wenn das KI für eine standardisierte Mittelwertdifferenz die Grenze 0,5 oder  $-0,5$  umfasst. In diesen Fällen, in denen ein mittelgroßer Effekt und gleichzeitig auch der Nulleffekt überdeckt wird, sollte eine Abwertung wegen fehlender Genauigkeit um 1 Stufe in Betracht gezogen werden. Eine Abwertung um 2 Stufen wird in Fällen in Betracht gezogen, in denen die zuvor genannten KI-Grenzen beide, d. h. z. B. im Fall binärer Daten sowohl 0,5 als auch 2 im KI enthalten sind.

#### **4.3.1.4.2 Aufwertung der Qualität der Evidenz**

##### **A: Große Effekte**

Die Qualität der Evidenz wird in methodisch guten, nicht randomisierten Interventionsstudien bei großen bzw. sehr großen Effekten um 1 oder 2 Stufen aufgewertet. Eine Aufwertung um 1

Stufe ist bei großen Effekten möglich, wenn beispielsweise das relative Risiko zwischen 2 und 5 bzw. zwischen 0,2 und 0,5 liegt, ohne dass plausible Confounder diesen Effekt verringern könnten. Eine Aufwertung um 1 weitere Stufe kann bei sehr großen Effekten erfolgen, wenn beispielsweise das relative Risiko über 5 bzw. unter 0,2 liegt, keine relevanten Verzerrungsrisiken vorliegen und die Effekte präzise (mit hinreichend schmalen KIs) geschätzt werden konnten [24].

### **B: Dosis-Wirkungs-Beziehung**

Die Qualität der Evidenz wird bei einer Dosis-Wirkungs-Beziehung aufgewertet [24].

### **C: Confounding**

Die Qualität der Evidenz wird bei einer möglichen Verstärkung des beobachteten Ergebnisses unter Berücksichtigung von potenziellen Confoundern oder Verzerrungsquellen aufgewertet [24].

## **4.3.2 Informationsdarstellung und Synthese von Studien zur diagnostischen Güte**

### **4.3.2.1 Darstellung der Einzelstudien**

Alle für den Evidenzbericht notwendigen Informationen werden aus den Unterlagen zu den für die Evidenzdarstellung berücksichtigten Studien in standardisierte Tabellen extrahiert. Ergeben sich im Abgleich der Informationen aus unterschiedlichen Dokumenten zu einer Studie (aber auch aus multiplen Angaben zu einem Aspekt innerhalb eines Dokumentes selbst) Diskrepanzen, die auf die Interpretation der Ergebnisse erheblichen Einfluss haben könnten, wird dies an den entsprechenden Stellen des Berichts dargestellt.

Die Extraktion erfolgt durch 1 Person und wird von einer 2. Person auf Grundlage der Studien kontrolliert.

Die Ergebnisse zu den in den Studien berichteten Vierfeldertafel-Daten werden im Bericht in Forestplots dargestellt. Als Zielgrößen für die diagnostische Güte werden primär Sensitivität und Spezifität herangezogen. Mithilfe der Gesamtschätzung von Sensitivität und Spezifität werden unter Annahme verschiedener Prävalenz-Szenarien Schätzungen für die Einträge einer Vierfeldertafel basierend auf 1000 untersuchte Personen angegeben. Zur Abschätzung der quantitativen Ergebnisunsicherheit werden ergänzend die oberen und unteren KI-Grenzen der Vierfeldertafel-Einträge basierend auf den KI-Grenzen der Gesamtschätzungen von Sensitivität und Spezifität und der angenommenen Prävalenz berechnet.

(Zu diesem Vorgehen gab es eine Spezifizierung im Projektverlauf, siehe Abschnitt 3.2.)

Das Vorgehen zur Bewertung des Einflusses des Verzerrungspotenzials auf die berichtsrelevanten Ergebnisse wird in Abschnitt 4.3.2.2 pro Studie beschrieben. Nach dieser

Bewertung werden die Informationen zur diagnostischen Güte in Evidenzprofilen [12] zusammengeführt und analysiert. Wenn möglich werden über die Darstellung der Ergebnisse der Einzelstudien hinaus die im Abschnitt 4.3.2.3 beschriebenen Verfahren eingesetzt.

Ergebnisse können ggf. im Evidenzbericht unberücksichtigt bleiben, wenn ein großer Anteil der in die Auswertung eigentlich einzuschließenden Personen nicht in der Auswertung berücksichtigt worden ist. Für die Entscheidung hierüber wird sich an einem Anteil von ca. 70 % orientiert, der in der Auswertung mindestens berücksichtigt sein sollte.

#### **4.3.2.2 Bewertung der Studienlimitationen (Verzerrungspotenzial)**

Das Verzerrungspotenzial der Ergebnisse wird zunächst pro Studie insbesondere anhand der nachfolgend aufgeführten Quality-Assessment-of-Diagnostic-Accuracy-Studies(QUADAS)-2-Kriterien [25,26] bewertet. Dazu erfolgt jeweils eine Bewertung mit „niedrig“, „unklar“ oder „hoch“. Die Voraussetzungen für eine Bewertung mit „niedrig“ werden im Folgenden erläutert. Eine Bewertung mit „unklar“ erfolgt grundsätzlich dann, wenn keine bzw. keine ausreichenden Angaben zur Bewertung zur Verfügung stehen.

- **Selektion der Patientinnen und Patienten**

Voraussetzung für eine Bewertung mit „niedrig“ ist, dass aus der Beschreibung der Studie hervorgeht, dass eine konsekutive oder zufällige Stichprobe von Patientinnen und Patienten eingeschlossen wurde und inadäquate Studienausschlüsse vermieden wurden.

- **Durchführung und Auswertung des Indextests**

Voraussetzung für eine Bewertung mit „niedrig“ ist, dass aus der Beschreibung der Studie hervorgeht, dass der Indextest ohne Wissen über die Ergebnisse des Referenztests ausgewertet wurde und bei Verwendung eines Grenzwerts dieser prospektiv festgelegt worden war.

- **Auswahl, Durchführung und Auswertung des Referenztests**

Voraussetzung für eine Bewertung mit „niedrig“ ist, dass aus der Beschreibung der Studie hervorgeht, dass der Referenztest die Zielerkrankung korrekt klassifiziert und ohne Wissen über die Ergebnisse des Indextests ausgewertet wurde.

- **Patientinnen- und Patientenfluss und zeitlicher Ablauf**

Voraussetzung für eine Bewertung mit „niedrig“ ist, dass aus der Beschreibung der Studie hervorgeht, dass die Zeitspanne zwischen Indextest und Referenztest adäquat war, alle Patientinnen und Patienten einen Referenzstandard erhalten haben, alle Patientinnen und Patienten denselben Referenzstandard erhalten haben und alle Patientinnen und Patienten in die Analyse einbezogen wurden.

Im Evidenzprofil werden abschließend auf Basis der Bewertung des Verzerrungspotenzials die Studienlimitationen studienübergreifend und zielgrößenspezifisch durch die Bewertungen „keine“, „schwerwiegend“ und „sehr schwerwiegend“ dargestellt.

#### **4.3.2.3 Metaanalysen**

Die Punktschätzungen und dazugehörigen univariaten 95 %-KI [27] aus den Studien werden mittels Forest Plots zusammenfassend dargestellt. Außerdem wird, sofern die dafür nötigen Anforderungen erfüllt sind, für die Testgütekriterien Sensitivität und Spezifität eine bivariate Metaanalyse durchgeführt [28]. Die Schätzung der Modellparameter erfolgt über ein generalisiertes lineares gemischtes Modell [29,30].

Falls die bivariate Metaanalyse präzise Schätzungen liefert, so werden bei Studien zur diagnostischen Güte die beobachteten Paare aus Sensitivität und Spezifität zweidimensional grafisch dargestellt. Des Weiteren werden die aus der bivariaten Metaanalyse gewonnenen Schätzungen für die Erwartungswerte als gepoolte Paare der Sensitivität und der Spezifität mit den dazugehörigen 95 %-Konfidenzregionen dargestellt [31].

Der Algorithmus zum Schätzen der Parameter im bivariaten Modell kann zu unpräzisen Schätzungen führen, das heißt zu Schätzungen mit zu großen Standardfehlern und entsprechenden Konfidenzregionen. Auch kann der Algorithmus gegebenenfalls keine Schätzungen liefern, wenn das Maximum-Likelihood-Verfahren nicht konvergiert. In beiden Fällen fehlen brauchbare Schätzungen. Die Gründe hierfür können beispielsweise sein, dass zu wenige Studien vorliegen oder dass einzelne Studien extreme Werte aufweisen. Sind die resultierenden Schätzungen unpräzise, werden die Ergebnisse der bivariaten Metaanalysen in der Regel nicht dargestellt. In diesem Fall wird auf univariate Metaanalysen für Sensitivität und Spezifität auf Basis von generalisierten gemischten Modellen zurückgegriffen. Von der Berechnung einer gepoolten Schätzung wird abgesehen, falls sich die KIs der eingehenden Studien nur gering oder gar nicht überlappen und gleichzeitig sich die Schätzungen der Studien zu stark unterscheiden.

#### **4.3.2.4 Bewertung der Qualität der Evidenz**

Alle für den Evidenzbericht relevanten Ergebnisse werden hinsichtlich einer Beeinflussung durch Studienlimitationen, des Risikos von Publikationsbias, der Genauigkeit und Konsistenz der Effekte, der Übertragbarkeit und ggf. hinsichtlich möglicher Aspekte zur Aufwertung der Qualität der Evidenz überprüft. Für die jeweiligen diagnostischen Eigenschaften wird eine studienübergreifende Aussage zur Qualität der Evidenz bezüglich des jeweiligen Vertrauens in die Schätzung der diagnostischen Güte getroffen. Hierzu erfolgt eine Einteilung der Qualität der Evidenz entsprechend der 4 Stufen der GRADE-Guideline in hoch, moderat, niedrig und sehr niedrig [18,19]:

- Das Vertrauen in die Schätzung der diagnostischen Güte ist hoch: Eine hohe Qualität der Evidenz bedeutet, dass der wahre Wert sehr sicher nahe bei der Schätzung liegt.
- Das Vertrauen in die Schätzung der diagnostischen Güte ist mäßig: Eine moderate Qualität der Evidenz bedeutet, dass der wahre Wert wahrscheinlich nahe bei der Schätzung liegt, aber die Möglichkeit besteht, dass er relevant verschieden ist.
- Das Vertrauen in die Schätzung der diagnostischen Güte ist begrenzt: Eine niedrige Qualität der Evidenz bedeutet, dass der wahre Wert durchaus relevant verschieden zur Schätzung sein kann.
- Das Vertrauen in die Schätzung der diagnostischen Güte ist sehr gering: Eine sehr niedrige Qualität der Evidenz bedeutet, dass der wahre Wert wahrscheinlich relevant verschieden von der Schätzung ist.

Die Bewertung erfolgt durch 2 Personen unabhängig voneinander. Diskrepanzen werden durch Diskussion zwischen den beiden aufgelöst.

Ergebnissen aus prospektiv geplanten Querschnitts- oder Kohortenstudien zur diagnostischen Güte mit einem adäquaten Referenztest wird im 1. Bewertungsschritt eine hohe Qualität der Evidenz attestiert [32].

#### **4.3.2.4.1 Abwertung der Qualität der Evidenz**

##### **A: Studienlimitationen**

Die Qualität der Evidenz wird aufgrund von schwerwiegenden oder sehr schwerwiegenden Studienlimitationen in der Regel um 1 oder 2 Stufen abgewertet, wenn starke Limitierungen in einem oder mehreren der in Abschnitt 4.3.2.2 genannten Kriterien das Vertrauen in die Schätzung der diagnostischen Güte beeinträchtigen. Die studienübergreifende Bewertung der Studienlimitationen erfolgt unter Berücksichtigung des Einflusses der einzelnen Studien auf die Schätzung der diagnostischen Güte [32].

##### **B: Inkonsistente (heterogene) Schätzer**

Die Qualität der Evidenz wird bei Inkonsistenz (unerklärter Heterogenität) zwischen Studienergebnissen in der Regel um 1 oder 2 Stufen abgewertet. Die Einschätzung einer möglichen Heterogenität erfolgt unter Berücksichtigung der Ähnlichkeit der Punktschätzungen und der Überlappung der KIs [33].

##### **C: Indirektheit**

Die Qualität der Evidenz wird bei starker oder sehr starker Indirektheit (eingeschränkter Übertragbarkeit) in der Regel um 1 oder 2 Stufen abgewertet. Indirektheit kann auf Abweichungen zwischen den Einschlusskriterien E1 (Population), E2 (Indextest) und E3

(Referenztest) in Tabelle 2 und den diesbezüglichen Einschlusskriterien der Studien basieren [32].

#### **D: Publikationsbias**

Die Qualität der Evidenz wird in der Regel um 1 Stufe abgewertet, wenn ein Publikationsbias wahrscheinlich oder sehr wahrscheinlich ist.

Kriterien für die Annahme eines Publikationsbias umfassen eine Evidenzgrundlage, welche vorrangig auf kleinen Studien und wenigen Ereignissen beruht, welche einen beobachteten Zusammenhang zwischen der Größe der Schätzung der diagnostischen Güte und der Studiengröße (oder Genauigkeit der Schätzung der diagnostischen Güte) zeigt oder welche Auffälligkeiten in der gesichteten Datenlage erkennen lässt [33].

#### **E: Fehlende Genauigkeit der Zielgrößen**

Die Qualität der Evidenz wird wegen geringer oder sehr geringer Genauigkeit der Schätzung in der Regel um 1 oder 2 Stufen abgewertet. Maßgeblich hierfür ist die Lage und Breite des 95 %-KI. Außerdem können sehr kleine Personenzahlen zu einer Abwertung wegen fehlender Genauigkeit führen [23].

(Zu diesem Vorgehen gab es eine Spezifizierung im Projektverlauf, siehe Abschnitt 3.2.)

#### **4.3.2.4.2 Aufwertung der Qualität der Evidenz**

In der Literatur gibt es bislang keine voll entwickelte Methodik zur Aufwertung der Qualität der Evidenz bei Studien zur diagnostischen Güte und es wird auf die Notwendigkeit einer weiteren theoretischen und empirischen Erarbeitung von Kriterien hingewiesen. Diskutiert wurden z. B. ein starker, konsistenter Zusammenhang von Sensitivität und Spezifität in der Receiver Operating Characteristic- (ROC)-Kurve oder sehr hohe Schätzungen für die diagnostische Güte, die nicht durch übrige Faktoren infrage gestellt werden [32,33].

Die genannten Faktoren werden, soweit möglich, betrachtet und eine mögliche Aufwertung im Einzelfall erwogen.

## **5 Ergebnisse**

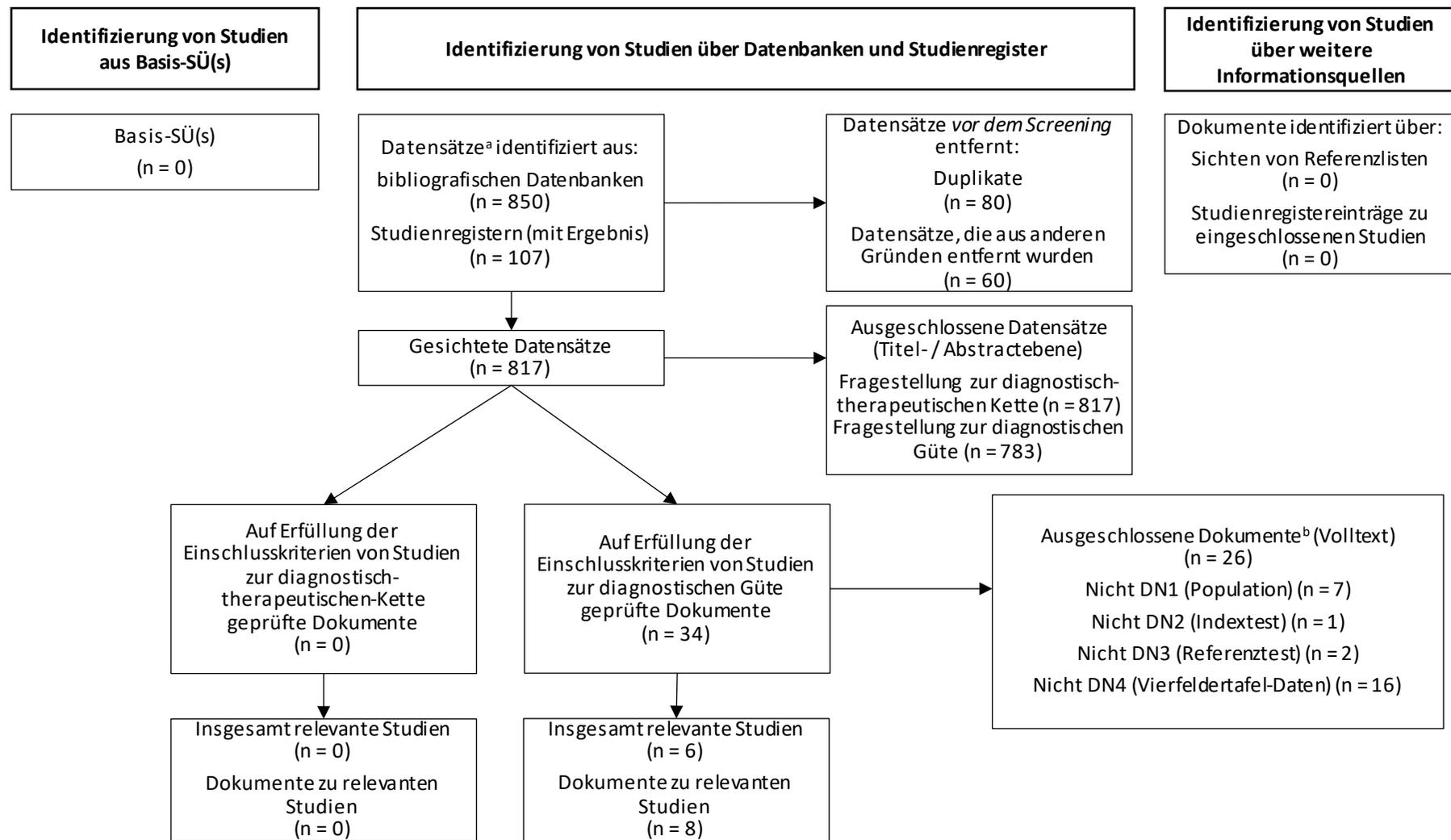
### **5.1 Ergebnisse der Informationsbeschaffung**

Abbildung 1 zeigt das Ergebnis der Informationsbeschaffung inklusive Studienselektion gemäß den Kriterien für den Studieneinschluss. Die Suchstrategien für die Suche in bibliografischen Datenbanken (letzte Suche am 06.04.2023) und die Suche in Studienregistern (letzte Suche am 27.04.2023) sind in Anhang C dargestellt.

Die Referenzen der als Volltexte geprüften, jedoch ausgeschlossenen Treffer finden sich mit Angabe des jeweiligen Ausschlussgrundes in Abschnitt 8.2.

Details zu den gestellten Autorenanfragen finden sich in Anhang B.

Die 1 identifizierte SÜ (siehe Abschnitt 8.1) wurde nicht als Basis-SÜ herangezogen.



a. Datensatz: Titel oder Abstract (oder beides) einer Studie, der in einer Datenbank (z. B. MEDLINE) oder auf einer Website indiziert ist.

b. nicht gelistete Ein- und Ausschlusskriterien (n = 0)

Abbildung 1: Ergebnis der Informationsbeschaffung

## Resultierender Studienpool

Durch die verschiedenen Rechenschritte wurden insgesamt 6 relevante Studien zur diagnostischen Güte identifiziert (siehe auch Tabelle 3). Es wurden keine Studien zur therapeutisch-diagnostischen-Kette identifiziert.

Tabelle 3: Studienpool der eingeschlossenen Studien zur diagnostischen Güte

Vergleich Studie	Verfügbare Dokumente	
	Vollpublikation (in Fachzeitschriften)	Registereintrag / Ergebnisbericht aus Studienregistern
<b>Wasserschlucktest im Vergleich zu instrumenteller Diagnostik</b>		
Hey 2023	ja [34,35]	nein
Horii 2021	ja [36,37]	nein
Patterson 2011	ja [38]	nein
<b>Klinische Schluckuntersuchung im Vergleich zu instrumenteller Diagnostik</b>		
Carnaby 2014	ja [39]	nein
Hey 2013	ja [40]	nein
Omori 2022 <sup>a</sup>	ja [41]	nein
a. Die Ergebnisse dieser Studie sind nicht verwertbar, da widersprüchliche Angaben zur Anzahl der Patientinnen und Patienten ohne Dysphagie und zur Anzahl der Patientinnen und Patienten mit Aspiration gegeben werden. Die Diskrepanz wurde in einer Autorenanfrage angesprochen, von der Autorengruppe jedoch nicht aufgelöst (Anhang B).		

1 der 6 eingeschlossenen Studien erfüllte zwar die formalen Einschlusskriterien, jedoch wurden die Daten für den Evidenzbericht nicht herangezogen. Daher wurde die Studie in der weiteren Evidenzdarstellung nicht berücksichtigt.

## 5.2 Charakteristika der für die Evidenzdarstellung berücksichtigten Studien

In Tabelle 4 werden die Charakteristika der für die Evidenzdarstellung berücksichtigten Studien dargestellt. Die in den jeweiligen Studien untersuchten Index- und Referenztests werden in Tabelle 5 charakterisiert. Die wesentlichen Ein- und Ausschlusskriterien für Patientinnen und Patienten in den berücksichtigten Studien sowie die Charakteristika der jeweiligen Studienpopulationen sind in Tabelle 6 und Tabelle 7 aufgeführt.

Tabelle 4: Charakteristika der für die Evidenzdarstellung berücksichtigten Studien zur diagnostischen Güte

Vergleich Studie	Studiendesign		Patienten- zahl N	Indextest	Referenztest: Untersuchungs- methode (Bewertungskala)	Ort und Zeitraum der Durchführung
	Einschluss der Patientinnen und Patienten <sup>a</sup>	Festlegung des Indextests				
<b>Wasserschlucktest im Vergleich zu instrumenteller Diagnostik</b>						
Hey 2023	prospektiv, unklar	I <sub>1</sub> : prospektiv I <sub>2</sub> : retrospektiv	203	I <sub>1</sub> : 51 ml WST I <sub>2</sub> : FraMaDySc	FEES (PAS / FOIS)	Deutschland, 2010 bis 2022
Horii 2021	prospektiv, unklar	retrospektiv	36	modifizierter WST <sup>b</sup>	VFSS (PAS)	Japan, 2013 bis 2016
Patterson 2011	prospektiv, konsekutiv	prospektiv	173	100 ml WST	FEES (k. A.)	England, über 25 Monate
<b>klinische Schluckuntersuchung im Vergleich zu instrumenteller Diagnostik</b>						
Carnaby 2014	prospektiv, unklar	retrospektiv	58 <sup>c</sup>	MASA-C	VFSS (Mann 2000 [42])	USA <sup>d</sup> , k. A.
Hey 2013	prospektiv, unklar	retrospektiv	80	klinische Schluck- untersuchung <sup>e</sup>	FEES (PAS / FOIS)	Deutschland, 2010 bis 2012
<p>a. Bewertet wird, ob der Einschluss der Patientinnen und Patienten prospektiv bzw. retrospektiv und konsekutiv bzw. nicht konsekutiv erfolgte.</p> <p>b. Der modifizierte WST war Teil eines Dysphagie-Screening-Tests und erfolgte nach negativem Testergebnis im repetitiven Speichelschlucktest und vor dem Test mit Essen und dem 30 ml WST. 28 Patientinnen und Patienten wurden mit dem modifizierten WST getestet. Die Ergebnisse zur diagnostischen Güte des 30 ml WST waren nicht verwertbar, da weniger als 50 % der eingeschlossenen Patientinnen und Patienten an diesem Test teilgenommen haben.</p> <p>c. Daten von 58 Patientinnen und Patienten konnten für die Studie genutzt werden. Es gibt keine Angaben zur Anzahl rekrutierter Patientinnen und Patienten.</p> <p>d. keine Angaben zum Ort der Durchführung, Ortsangabe bezieht sich auf die Affiliation der Autorengruppe</p> <p>e. In der Studie wurde die diagnostische Güte von 6 Indextests berichtet. In diesem Evidenzbericht werden die 2 Indextests berichtet, die mehrere für die Schluckfunktion relevante Aspekte (<math>\geq 3</math> klinische Parameter und <math>\geq 2</math> klinische Parameter) untersuchen (siehe auch Tabelle 5).</p> <p>FEES: fiberoptisch endoskopische Evaluation des Schluckvorganges; FOIS: Functional Oral Intake Scale; FraMaDySc: Frankfurt/Marburg Dysphagie Screening; I: Indextest; k. A.: keine Angabe; MASA-C: Mann Assessment of Swallowing Ability-Cancer; N: Anzahl eingeschlossener Patientinnen und Patienten; PAS: Penetration-Aspiration-Skala; VFSS: Videofluoroskopische Schluckuntersuchung; WST: Wasserschlucktest</p>						

Tabelle 5 Charakterisierung der Index- und Referenztests in den berücksichtigten Studien zur diagnostischen Güte (mehrsseitige Tabelle)

Vergleich Studie	Indextest	Referenztest
<b>Wasserschlucktest im Vergleich zu instrumenteller Diagnostik</b>		
Hey 2023	<p><b>Test(s):</b> 51 ml WST (I<sub>1</sub>) bzw. FraMaDySc (I<sub>2</sub>)</p> <p><u>Testzeitpunkt:</u></p> <ul style="list-style-type: none"> <li>▪ postoperativ</li> <li>▪ nach Zustimmung der Chirurgin / des Chirurgen</li> <li>▪ vor der ersten oralen Nahrungsaufnahme</li> </ul> <p><u>Ablauf Schluckversuche mit Wasser:</u></p> <ul style="list-style-type: none"> <li>▪ 2 x<sup>a</sup> je 2 ml per Löffel</li> <li>▪ 2 x<sup>a</sup> je 5 ml per Becher</li> <li>▪ je 1 x 10 ml bzw. 20 ml per Becher</li> <li>▪ Gesamtmenge: 51 ml, wenn die Schluckversuche mit 2 ml und 5 ml jeweils 3 x durchgeführt werden</li> </ul> <p><u>Trennwert:</u> Testpositiv bei Vorliegen von einem der folgenden Kriterien:</p> <ul style="list-style-type: none"> <li>▪ WST bzw. FraMaDySc: gurgelige Stimmgebung<sup>b</sup> ohne erfolgreichen Reinigungsversuch vor dem ersten Schluckversuch mit 2 ml</li> <li>▪ WST: gurgelige Stimmgebung<sup>b</sup> vor einem Schluckversuch mit 5, 10 oder 20 ml</li> <li>▪ WST bzw. FraMaDySc: gurgelige Stimmgebung<sup>b</sup> oder veränderte Stimmgebung nach dem Schlucken</li> <li>▪ WST bzw. FraMaDySc: Husten oder Räuspern nach dem Schlucken</li> </ul> <p><u>Testabbruch:</u> Testpositiv nach Schluckversuch mit 2 ml, 5 ml, 10 ml oder 20 ml</p> <p><u>Anzahl Testdurchführende mit Testerfahrung:</u> 7 Sprachtherapeutinnen oder -therapeuten bzw. Logopädinnen oder Logopäden (geschult in der Durchführung des WST; mehrheitlich mindestens 5-jährige therapeutische Erfahrung mit Schluckstörungen)</p>	<p><b>Test(s):</b> FEES mit Bewertung der Dysphagie, Aspiration und Oralisierungseinschränkung nach PAS und FOIS</p> <p><u>Testzeitpunkt:</u></p> <ul style="list-style-type: none"> <li>▪ innerhalb einer Stunde nach Beendigung des WST</li> </ul> <p><u>Ablauf Schluckversuche mit verschiedenen Konsistenzen (Menge):</u></p> <ul style="list-style-type: none"> <li>▪ Wasser (je 2, 5, 10 und 20 ml)</li> <li>▪ Püree (2 und 5 ml)</li> <li>▪ Keks (ein Bissen)</li> </ul> <p><u>Trennwert für Referenzkriterien:</u></p> <ul style="list-style-type: none"> <li>▪ Dysphagie: PAS ≥ 4 und FOIS ≤ 4</li> <li>▪ Aspiration: PAS ≥ 6</li> <li>▪ Oralisierungseinschränkung: FOIS ≤ 4</li> </ul> <p><u>Testabbruch:</u> bei Auftreten einer Aspiration ohne Verbesserung der Schluckfunktion durch therapeutisches Manöver</p> <p><u>Anzahl Testdurchführende (Erfahrung):</u> 2 Phoniaterinnen oder Phoniater (mindestens 10 Jahre Erfahrung mit FEES)</p>

Tabelle 5 Charakterisierung der Index- und Referenztests in den berücksichtigten Studien zur diagnostischen Güte (mehrsseitige Tabelle)

Vergleich Studie	Indextest	Referenztest
Horii 2021	<p><b>Test(s):</b> modifizierter WST</p> <p><u>Testzeitpunkt:</u></p> <ul style="list-style-type: none"> <li>▪ postoperativ</li> <li>▪ nach negativem Testergebnis beim wiederholtem Speichelschlucktest</li> </ul> <p><u>Ablauf Schluckversuche mit Wasser:</u></p> <ul style="list-style-type: none"> <li>▪ 1 bis 3 x 3 ml Wasser</li> </ul> <p><u>Trennwert:</u> Testpositiv bei Vorliegen von einem der ersten 3 Kriterien (Score &lt; 4):</p> <p>(1) erfolgloses Schlucken mit Husten und / oder veränderter Atmung<sup>c</sup></p> <p>(2) erfolgreiches Schlucken ohne Husten aber mit veränderter Atmung oder gurgeliger Stimmgebung<sup>d</sup></p> <p>(3) erfolgreiches Schlucken mit Husten und/ oder gurgeliger Stimmgebung<sup>e</sup>.</p> <p>(4) erfolgreiches Schlucken ohne Husten und ohne gurgeliger Stimmgebung<sup>f</sup></p> <p>(5) 2 erfolgreiche Schluckversuche innerhalb von 30 Sekunden ohne Husten und ohne gurgeliger Stimmgebung<sup>g</sup></p> <p><u>Testabbruch:</u> Subjektive Einschätzung der testdurchführenden Person eines erhöhten Aspirationsrisikos, welches bei weiteren Schluckversuchen zu einer Patientengefährdung führen könnte.</p> <p><u>Anzahl Testdurchführende (Erfahrung):</u> k. A.</p>	<p><b>Test(s):</b> VFSS mit Bewertung der Aspiration nach PAS</p> <p><u>Testzeitpunkt:</u></p> <ul style="list-style-type: none"> <li>▪ postoperativ</li> <li>▪ nach Zustimmung der behandelnden Ärztin oder des behandelnden Arztes</li> <li>▪ k. A. zu welchem Zeitpunkt der Referenztest im Vergleich zum Indextest durchgeführt wurde</li> </ul> <p><u>Ablauf Schluckversuche mit verschiedenen Konsistenzen (Menge):</u></p> <ul style="list-style-type: none"> <li>▪ honigartig angedicktes Wasser (3 ml)</li> <li>▪ nektarartig angedicktes Wasser (3 ml)</li> <li>▪ Wasser (3 ml)</li> </ul> <p><u>Trennwert für Referenzkriterium:</u></p> <ul style="list-style-type: none"> <li>▪ erhöhtes Aspirationsrisiko (unsicheres Schlucken): PAS ≥ 3</li> </ul> <p><u>Testabbruch:</u> bei Aspiration oder bei erhöhtem Aspirationsrisiko, welches durch die durchführende Ärztin oder den durchführenden Arzt eingeschätzt wird.</p> <p><u>Anzahl Testdurchführende (Erfahrung):</u> k. A.</p>

Tabelle 5 Charakterisierung der Index- und Referenztests in den berücksichtigten Studien zur diagnostischen Güte (mehreseitige Tabelle)

Vergleich Studie	Indextest	Referenztest
Patterson 2011	<p><b>Test(s):</b> 100 ml WST</p> <p><u>Testzeitpunkt:</u></p> <ul style="list-style-type: none"> <li>▪ 3 Monate nach der Tumorbehandlung</li> </ul> <p><u>Ablauf Schluckversuche mit Wasser:</u></p> <ul style="list-style-type: none"> <li>▪ Patientinnen und Patienten wurden aufgefordert, schnellstmöglich 100 ml Wasser zu trinken</li> </ul> <p><u>Trennwert:</u></p> <ul style="list-style-type: none"> <li>▪ Husten oder Würgen nach dem Schlucken</li> <li>▪ gurgelige Stimmgebung<sup>b</sup> nach dem Schlucken</li> <li>▪ Unvermögen, die Gesamtmenge zu trinken</li> </ul> <p><u>Testabbruch:</u></p> <ul style="list-style-type: none"> <li>▪ Patientenangabe von Husten bei Flüssigkeiten</li> <li>▪ Husten während des Schluckversuchs</li> </ul> <p><u>Anzahl Testdurchführende (Erfahrung):</u> 1 Sprachtherapeutin (k. A.)</p>	<p><b>Test(s):</b> FEES mit Bewertung der Aspiration</p> <p><u>Testzeitpunkt:</u></p> <ul style="list-style-type: none"> <li>▪ siehe Indextest</li> <li>▪ k. A., zu welchem Zeitpunkt der Referenztest im Vergleich zum Indextest durchgeführt wurde</li> </ul> <p><u>Ablauf Schluckversuche mit verschiedenen Konsistenzen (Menge):</u></p> <ul style="list-style-type: none"> <li>▪ blau eingefärbte Milch (10 ml)</li> </ul> <p><u>Trennwert für Referenzkriterium:</u></p> <ul style="list-style-type: none"> <li>▪ Vorliegen einer Aspiration ermittelt durch Positionierung des Endoskops posterior zur Epiglottis (k. A. zur weiteren Definition)</li> </ul> <p><u>Testabbruch:</u></p> <ul style="list-style-type: none"> <li>▪ bei Aspiration</li> </ul> <p><u>Anzahl Testdurchführende (Erfahrung):</u> 1 Sprachtherapeutin (k. A.)</p>

Tabelle 5 Charakterisierung der Index- und Referenztests in den berücksichtigten Studien zur diagnostischen Güte (mehrsseitige Tabelle)

Vergleich Studie	Indextest	Referenztest
<b>klinische Schluckuntersuchung im Vergleich zu instrumenteller Diagnostik</b>		
Carnaby 2014	<p><b>Test(s):</b> Mann Assessment of Swallowing Ability – Cancer (MASA-C)</p> <p><u>Testzeitpunkt:</u></p> <ul style="list-style-type: none"> <li>nach Beendigung der Tumorbehandlung</li> </ul> <p><u>Ablauf MASA-C:</u></p> <ul style="list-style-type: none"> <li>überprüft werden 24 für die Schluckfunktion relevante Aspekte<sup>h</sup> (davon wurden 15 aus dem MASA übernommen und 9 krebsspezifische ergänzt)</li> <li>MASA-C Gesamtpunktzahl: 200</li> </ul> <p><u>Trennwert:</u></p> <ul style="list-style-type: none"> <li>Dysphagie: ≤ 185 Punkte</li> <li>Aspiration: ≤ 176 Punkte</li> </ul> <p><u>Testabbruch:</u> k. A.</p> <p><u>Anzahl Testdurchführende (Erfahrung):</u> 1 Sprachtherapeutin und 1 Sprachtherapeut (k. A.)</p>	<p><b>Test(s):</b> VFSS mit Bewertung der Dysphagie, nach den Kriterien von Mann 2000</p> <p><u>Testzeitpunkt:</u></p> <ul style="list-style-type: none"> <li>siehe Indextest</li> <li>VFSS wurde am gleichen Tag wie der Indextest durchgeführt.</li> </ul> <p><u>Ablauf Schluckversuche mit verschiedenen Konsistenzen (Menge):</u></p> <ul style="list-style-type: none"> <li>Flüssigkeit (5 und 10 ml)</li> <li>nektarartig angedickte Flüssigkeit (5 und 10 ml)</li> <li>Pudding (5 und 10 ml)</li> </ul> <p><u>Trennwert für Referenzkriterium:</u></p> <ul style="list-style-type: none"> <li>Vorliegen einer Dysphagie (mild bis komplett) nach den Kriterien von Mann 2000 [42]</li> </ul> <p><u>Testabbruch:</u> k. A.</p> <p><u>Anzahl Testdurchführende (Erfahrung):</u> 1 Sprachtherapeutin oder -therapeut und 1 Radiologin bzw. Radiologe (k. A.)</p>

Tabelle 5 Charakterisierung der Index- und Referenztests in den berücksichtigten Studien zur diagnostischen Güte (mehrsseitige Tabelle)

Vergleich Studie	Indextest	Referenztest
Hey 2013	<p><b>Test(s):</b> Klinische Schluckuntersuchung</p> <p><u>Testzeitpunkt:</u></p> <ul style="list-style-type: none"> <li>▪ Nach der Tumorentfernung mit Zustimmung der Chirurgin / des Chirurgen zur orale (Nahrungs-) Aufnahme</li> <li>▪ nach Beendigung der Tumorbehandlung</li> </ul> <p><u>Ablauf klinische Schluckuntersuchung:</u></p> <ul style="list-style-type: none"> <li>▪ überprüft werden 7 für die Schluckfunktion relevante Aspekte: Sprachstörung, gurgelige Stimmgebung<sup>b</sup>, willentliches Husten, Mundöffnung, Beweglichkeit der Zunge, Zungenkraft und Würgreflex</li> </ul> <p><u>Trennwert:</u></p> <ul style="list-style-type: none"> <li>▪ Testpositiv: Bestehen oder Fehlen des getesteten Aspekts<sup>i</sup></li> </ul> <p><u>Testabbruch:</u> k. A.</p> <p><u>Anzahl Testdurchführende (Erfahrung):</u> 2 Sprachtherapeutinnen oder -therapeuten bzw. Logopädinnen oder Logopäden (≥ 5 Jahre Erfahrung in der Bewertung und Behandlung von Schluckstörungen)</p>	<p><b>Test(s):</b> FEES mit Bewertung der Dysphagie, Aspiration und Oralisierungseinschränkung nach PAS und FOIS</p> <p><u>Testzeitpunkt:</u> Innerhalb einer Stunde nach der klinischen Schluckuntersuchung</p> <p><u>Ablauf Schluckversuche mit verschiedenen Konsistenzen (Menge):</u></p> <ul style="list-style-type: none"> <li>▪ Wasser (2, 5, 10 und 20 ml)</li> <li>▪ pürierte Konsistenzen (k. A.)</li> <li>▪ feste Konsistenzen (k. A.)</li> </ul> <p><u>Trennwert für Referenzkriterien:</u></p> <ul style="list-style-type: none"> <li>▪ Dysphagie: PAS ≥ 4 und FOIS ≤ 4</li> <li>▪ Aspiration: PAS ≥ 6</li> <li>▪ Oralisierungseinschränkung: FOIS ≤ 4</li> </ul> <p><u>Testabbruch:</u> bei Auftreten einer Aspiration ohne Verbesserung der Schluckfunktion durch therapeutische Manöver</p> <p><u>Anzahl Testdurchführende (Erfahrung):</u> 1 Phoniaterin (mindestens 10 Jahre Erfahrung mit FEES)</p>
<p>a. Sofern einer der 2 Schluckversuche positiv ausfällt, erfolgt 1 dritter Schluckversuch. Bei positivem Ergebnis im dritten Schluckversuch erfolgt der Abbruch des WST.</p> <p>b. im Englischen: <i>wet voice</i></p> <p>c. im Englischen: failed swallowing with choking and/or changes in breathing</p> <p>d. im Englischen: successful swallowing without choking but with changes in breathing or wet hoarseness</p> <p>e. im Englischen: successful swallowing but with choking and/or wet hoarseness</p> <p>f. im Englischen: successful swallowing with no choking or wet hoarseness</p> <p>g. im Englischen: two successful swallows with no choking or wet hoarseness within 30s</p> <p>h. Der MASA-C wurde in Carnaby 2014 entwickelt. Die 24 Aspekte des MASA-C sind dort beschrieben.</p> <p>i. Da 3 der 7 Aspekte nicht mit den Skalen korrelierten, wurden die 4 Aspekte Sprachstörung, feucht gurgelige Stimmgebung, Beweglichkeit der Zunge und Zungenkraft als Trennwert untersucht.</p> <p>FEES: fiberoptisch endoskopische Evaluation des Schluckvorganges; FraMaDySc: Frankfurt/Marburg Dysphagie Screening; FOIS: Functional Oral Intake Scale; I: Indextest; k. A.: keine Angabe; MASA: Mann Assessment of Swallowing Ability; PAS: Penetration-Aspiration-Skala; VFSS: videofluoroskopische Schluckuntersuchung; WST: Wasserschlucktest</p>		

Tabelle 6: Ein- / Ausschlusskriterien für Patientinnen und Patienten in den berücksichtigten Studien zur diagnostischen Güte

Vergleich Studie	Wesentliche Einschlusskriterien	Wesentliche Ausschlusskriterien
<b>Wasserschlucktest im Vergleich zu instrumenteller Diagnostik</b>		
Hey 2023	<ul style="list-style-type: none"> <li>▪ Alter: ≥ 18 Jahre</li> <li>▪ Patientinnen und Patienten mit KHT (UICC Stadium II–IV)</li> <li>▪ nach Tumoroperation, einschließlich Resektion und, bei Bedarf, plastischer Rekonstruktion</li> <li>▪ Einwilligungserklärung</li> </ul>	<ul style="list-style-type: none"> <li>▪ Anamnese einer Schluckstörung</li> <li>▪ neurologische Erkrankungen (z. B. Parkinsonkrankheit)</li> <li>▪ totale Laryngektomie</li> </ul>
Horii 2021	<ul style="list-style-type: none"> <li>▪ Alter: k. A.</li> <li>▪ Patientinnen und Patienten mit KHT</li> <li>▪ nach Tumorsektion</li> </ul>	<ul style="list-style-type: none"> <li>▪ Anamnese einer Schluckstörung: k. A.</li> <li>▪ Kognitive Einschränkungen mit einem HDS-R Score ≤ 20 (z. B. aufgrund einer schweren Demenz)</li> <li>▪ Studienabbruch aufgrund Umzugs oder anderer Gründe</li> </ul>
Patterson 2011	<ul style="list-style-type: none"> <li>▪ Alter: k. A.</li> <li>▪ Patientinnen und Patienten mit Larynx-, Hypopharynx-, Oropharynx- oder Nasopharynxkarzinom oder mit unbekanntem Primärtumor</li> <li>▪ geplante kurative Strahlen- oder Strahlenchemotherapie</li> </ul>	<ul style="list-style-type: none"> <li>▪ Anamnese einer Schluckstörung</li> <li>▪ Rezidiv oder eine Residualerkrankung</li> </ul>
<b>klinische Schluckuntersuchung im Vergleich zu instrumenteller Diagnostik</b>		
Carnaby 2014	<ul style="list-style-type: none"> <li>▪ Alter: k. A.</li> <li>▪ Patientinnen und Patienten mit diagnostiziertem KHT im Oropharynx oder in angrenzender Region</li> <li>▪ geplante Strahlentherapie</li> </ul>	<ul style="list-style-type: none"> <li>▪ Anamnese einer nicht oralen Ernährung bei krebbsbedingter Erkrankung</li> <li>▪ Anamnese weiterer Schluckstörungen: k. A.</li> </ul>
Hey 2013	<ul style="list-style-type: none"> <li>▪ Alter: 18–99 Jahre</li> <li>▪ Patientinnen und Patienten mit KHT (UICC Stadium II–IV)</li> <li>▪ nach Tumoroperation</li> </ul>	<ul style="list-style-type: none"> <li>▪ Anamnese einer Schluckstörung</li> <li>▪ neurologische Erkrankungen</li> </ul>
HDS-R: Hasegawa's Dementia Scale, revised; k. A.: keine Angabe; KHT: Kopf-Hals-Tumor; UICC: Union for International Cancer Control		

Tabelle 7: Charakterisierung der Studienpopulationen in den berücksichtigten Studien zur diagnostischen Güte (mehreseitige Tabelle)

<b>Vergleich</b>	
<b>Studie</b>	
<b>Charakteristika</b>	
<b>Wasserschlucktest im Vergleich zu instrumenteller Diagnostik</b>	
<b>Hey 2023</b>	
N	184
Alter [< 70 Jahre / ≥ 70 Jahre] %	81 / 19
Geschlecht [w / m] %	29 / 71
Body-Mass-Index [kg/m <sup>2</sup> ] MW (SD)	k. A.
UICC Tumorstadium, n (%)	
II	46 (25)
III	30 (16)
IV	108 (59)
Tumorlokalisierung, n (%)	
Mundhöhle	67 (36)
Oropharynx	93 (51)
Hypopharynx / Larynx	24 (13)
<b>Horii 2021</b>	
N	36
Alter [Jahre] MW (Min; Max)	69,2 (38; 85)
Geschlecht [w / m] %	31 / 69
Body-Mass-Index [kg/m <sup>2</sup> ] MW (Min; Max)	21,4 (14,9; 26,9)
Tumorstadium, n (%)	k. A.
Tumorlokalisierung, n (%)	
Mundhöhle <sup>a</sup>	25 <sup>b</sup> (69) <sup>b</sup>
Larynx	1 (3)
Oropharynx	4 (11)
Hypopharynx	3 (8)
Ösophagus	2 (6)
zervikale Lymphknoten	1 (3)
Tumorbehandlung, n (%)	
Chemotherapie	7 (19)
Strahlentherapie	12 (33)
Rekonstruktion	25 (70)
Tracheotomie	22 (61)
Entfernung der Halslymphknoten [unilateral / bilateral]	22 / 7 (61 / 19)

Tabelle 7: Charakterisierung der Studienpopulationen in den berücksichtigten Studien zur diagnostischen Güte (mehreseitige Tabelle)

<b>Vergleich</b>	
<b>Studie</b>	
<b>Charakteristika</b>	
<b>Patterson 2011<sup>c</sup></b>	
N	110
Alter [Jahre] MW (SW)	62 (42-87)
Geschlecht [w / m] %	16 <sup>b</sup> / 84 <sup>b</sup>
Body-Mass-Index [kg/m <sup>2</sup> ] MW (SD)	k. A.
Tumorgröße, n (%) <sup>b, d</sup>	
T1-2	59 (54)
T3-4	39 (39)
Tumorlokalisierung, n (%) <sup>b</sup>	
Oropharynx	42 (38)
Hypopharynx	7 (6)
Larynx	43 (39)
Nasopharynx	6 (6)
Unbekannter Primärtumor	12 (11)
Tumorbehandlung, n (%) <sup>b</sup>	
Strahlentherapie 50/16 Gy	26 (24)
Strahlentherapie 63/30 Gy	19 (17)
Strahlenchemotherapie	65 (59)
<b>klinische Schluckuntersuchung im Vergleich zu instrumenteller Diagnostik</b>	
<b>Carnaby 2014</b>	
N	58
Alter [Jahre] MW (SD)	57,6 (11,35)
Geschlecht [w / m] n	14 / 44
Body-Mass-Index [kg/m <sup>2</sup> ] MW (SD)	28,8 (1,0)
Tumorgröße [T], Median (SW)	2 (0–4)
Tumorlokalisierung, n (%) <sup>b, e</sup>	
Zungengrund	11 (19)
Tonsillen	16 (28)
Tumorbehandlung, n (%) <sup>b</sup>	
zusätzliche Chemotherapie	22 (38)
Entfernung der Lymphknoten	22 (38)
Strahlentherapiedosis in Gy MW (SD)	69,7 (8,12)

Tabelle 7: Charakterisierung der Studienpopulationen in den berücksichtigten Studien zur diagnostischen Güte (mehreseitige Tabelle)

<b>Vergleich</b>	
<b>Studie</b>	
<b>Charakteristika</b>	
<b>Hey 2013</b>	
N	80
Alter [ $\leq$ 40 Jahre / 41-60 Jahre / $>$ 60 Jahre] n (%)	4 / 34 / 42 (5 / 42 / 53)
Geschlecht [w / m] %	27 / 73
Body-Mass-Index [ $\text{kg}/\text{m}^2$ ] MW (SD)	k. A.
UICC Tumorstadium, n (%)	
II	23 (29)
III	9 (11)
IV	48 (60)
Tumorlokalisation	
Mundhöhle	30 (37)
Oropharynx	34 (43)
Hypopharynx / Larynx	16 (20)
a. Zusammenfassung der Studienangaben zu Zahnfleisch (Ober- und Unterkiefer), Mundschleimhaut, Zunge und Mundboden.	
b. eigene Berechnung	
c. Charakteristika der Studienpopulation werden lediglich für die Patientinnen und Patienten angegeben, die an allen 4 Testzeitpunkten teilgenommen haben.	
d. Das Tumorstadium der übrigen 12 Patientinnen und Patienten ist unklar.	
e. Es wurden Patientinnen und Patienten mit oropharyngealem Tumor oder Tumor in angrenzenden Regionen eingeschlossen. Die genaue Lokalisation des Tumors der übrigen 31 Patientinnen und Patienten ist unklar.	
k. A.: keine Angabe; m: männlich; Max: Maximum; Min: Minimum; MW: Mittelwert; N: Anzahl der Patientinnen und Patienten mit Charakteristika zu Studienbeginn; n: Anzahl der Patientinnen und Patienten mit Merkmal; SD: Standardabweichung; SW: Spannweite; UICC: Union for International Cancer Control; T: Größe des Primär Tumors; w: weiblich	

### 5.3 Ergebnisse zur diagnostischen Güte

#### 5.3.1 Ergebnisse zur diagnostischen Güte: Referenzkriterium Dysphagie

##### 5.3.1.1 Ergebnisse zum Wasserschlucktest im Vergleich zu instrumenteller Diagnostik

Tabelle 8: Evidenzprofil zur diagnostischen Güte zu Dysphagie von Wasserschlucktest im Vergleich zu instrumenteller Diagnostik (mehrsseitige Tabelle)

N (n)	Vierfelder- tafel- Ergebnisse	Faktoren der Qualität der Evidenz						Testgüte % [95 %-KI]	Schätzung für 1000 untersuchte Personen [95 %-KI]			Qualität der Evidenz <sup>a</sup>
		Studien- limitationen <sup>b</sup>	Inkonsis- tenz	Indirekt- heit	Publikations- bias	Fehlende Genauig- keit	Andere Faktoren		Prävalenz- Annahme: 55 % <sup>c</sup>	Prävalenz- Annahme: 65 %	Prävalenz- Annahme: 75 %	
<b>51 ml WST</b>												
1 (184) [34]	RP	nicht schwer- wiegend	nicht zu- treffend	nicht schwer- wiegend	unentdeckt	schwer- wiegend <sup>d</sup>	keine	<b>Sensitivität:</b> 91 % [84 %; 95 %]	500 [463; 524] <sup>e</sup>	590 [547; 620]	681 [631; 715]	moderat
	FN							50 [26; 87] <sup>f</sup>	60 [30; 103]	69 [35; 119]		
	RN	nicht zu- treffend	schwer- wiegend <sup>g</sup>	<b>Spezifität:</b> 84 % [73 %; 92 %]	380 [329; 415] <sup>h</sup>	295 [256; 323]	211 [183; 231]	moderat				
	FP			70 [35; 121] <sup>i</sup>	55 [27; 94]	39 [19; 67]						

Tabelle 8: Evidenzprofil zur diagnostischen Güte zu Dysphagie von Wasserschlucktest im Vergleich zu instrumenteller Diagnostik (mehrsseitige Tabelle)

N (n)	Vierfelder- tafel- Ergebnisse	Faktoren der Qualität der Evidenz						Testgüte % [95 %-KI]	Schätzung für 1000 untersuchte Personen [95 %-KI]			Qualität der Evidenz <sup>a</sup>
		Studien- limitationen <sup>b</sup>	Inkonsis- tenz	Indirekt- heit	Publikations- bias	Fehlende Genauig- keit	Andere Faktoren		Prävalenz- Annahme: 55 % <sup>c</sup>	Prävalenz- Annahme: 65 %	Prävalenz- Annahme: 75 %	
<b>FraMaDySc</b>												
1 (184) [34]	<b>RP</b>	sehr schwer- wiegend <sup>d</sup>	nicht zu- treffend	nicht schwer- wiegend	unentdeckt	schwer- wiegend <sup>d</sup>	keine	<b>Sensitivität:</b> 91 % [84 %; 95 %]	500 [463; 524]	590 [547; 620]	681 [631; 715]	niedrig <sup>k</sup>
	<b>FN</b>										50 [26; 87]	
	<b>RN</b>	nicht zu- treffend		schwer- wiegend <sup>e</sup>	<b>Spezifität:</b> 88 % [77 %; 94 %]	394 [346; 425]	306 [269; 331]	219 [192; 236]	niedrig <sup>k</sup>			
	<b>FP</b>						56 [25; 104]	44 [19; 81]		31 [14; 58]		

Tabelle 8: Evidenzprofil zur diagnostischen Güte zu Dysphagie von Wasserschlucktest im Vergleich zu instrumenteller Diagnostik (mehrsseitige Tabelle)

N (n)	Vierfelder- tafel- Ergebnisse	Faktoren der Qualität der Evidenz						Testgüte % [95 %-KI]	Schätzung für 1000 untersuchte Personen [95 %-KI]			Qualität der Evidenz <sup>a</sup>
		Studien- limitationen <sup>b</sup>	Inkonsis- tenz	Indirekt- heit	Publikations- bias	Fehlende Genauig- keit	Andere Faktoren		Prävalenz- Annahme: 55 % <sup>c</sup>	Prävalenz- Annahme: 65 %	Prävalenz- Annahme: 75 %	
<p>a. Unter Qualität der Evidenz wird die studienübergreifende zielgrößenbezogene Qualität der Evidenz verstanden.</p> <p>b. Die Kriterien des Verzerrungspotenzials pro Studie sind Anhang A zu entnehmen.</p> <p>c. Die Festlegung der Prävalenz-Annahmen orientierte sich an den Prävalenzraten der eingeschlossenen Studien mit verwertbaren Daten. Exemplarische Erläuterung zur Interpretation: Eine Prävalenz-Annahme von 55 % bedeutet für 1000 untersuchte Personen, dass 550 Personen eine Dysphagie haben.</p> <p>d. Das 95 %-KI der Sensitivität überdeckt die Grenze von 90 %.</p> <p>e. Exemplarische Erläuterung zur Interpretation: Bei 463 bis 524 von 550 Personen wird richtigerweise eine Dysphagie diagnostiziert und die Personen erhalten eine instrumentelle Diagnostik.</p> <p>f. Exemplarische Erläuterung zur Interpretation: Bei 26 bis 87 von 550 Personen wird fälschlicherweise keine Dysphagie diagnostiziert und die Personen erhalten keine instrumentelle Diagnostik und keine Therapie.</p> <p>g. Das 95 %-KI der Spezifität überdeckt die Grenze von 80 %.</p> <p>h. Exemplarische Erläuterung zur Interpretation: Bei 329 bis 415 von 450 Personen wird richtigerweise diagnostiziert, dass sie keine Dysphagie haben und ihnen wird eine instrumentelle Diagnostik erspart.</p> <p>i. Exemplarische Erläuterung zur Interpretation: Bei 35 bis 121 von 450 Personen wird fälschlicherweise eine Dysphagie diagnostiziert und sie haben durch den Indextest einen weiteren Test zusätzlich zur instrumentellen Diagnostik.</p> <p>j. Das Kriterium Indextest wurde mit hoch bewertet.</p> <p>k. Aufgrund der wenigen Unterschiede im Testablauf und der gleichen bzw. ähnlichen diagnostischen Güte zum 51 ml Wasserschlucktest wurde die Qualität der Evidenz um 2 und nicht um 3 Stufen abgewertet.</p> <p>FN: falsch-negativ (Personen, bei denen fälschlicherweise keine Dysphagie diagnostiziert wurde); FP: falsch-positiv (Personen, bei denen fälschlicherweise eine Dysphagie diagnostiziert wurde); FraMaDySc: Frankfurt/Marburg Dysphagia Screening; KI: Konfidenzintervall; N: Anzahl der Studien; n: Anzahl ausgewerteter Personen; RN: richtig-negativ (Personen ohne Dysphagie); RP: richtig-positiv (Personen mit Dysphagie); WST: Wasserschlucktest</p>												

### 5.3.1.2 Ergebnisse zur klinischen Schluckuntersuchung im Vergleich zu instrumenteller Diagnostik

Tabelle 9: Evidenzprofil zur diagnostischen Güte zu Dysphagie von klinischer Schluckuntersuchung im Vergleich zu instrumenteller Diagnostik (mehrsseitige Tabelle)

N (n)	Vierfelder- tafel-Ergeb- nisse	Faktoren der Qualität der Evidenz						Testgüte % [95 %-KI]	Schätzung für 1000 untersuchte Personen [95 %-KI]			Qualität der Evidenz <sup>a</sup>	
		Studien- limitationen <sup>b</sup>	Inkonsis- tenz	Indirekt- heit	Publika- tions- bias	Fehlende Genauig- keit	Andere Faktoren		Prävalenz- Annahme: 55 % <sup>c</sup>	Prävalenz- Annahme: 65 %	Prävalenz- Annahme: 75 %		
<b>MASA-C</b>													
1 (50) [39]	RP	sehr schwer- wiegend <sup>d</sup>	nicht zu- treffend	nicht schwer- wiegend	unent- deckt	schwer- wiegend <sup>e</sup>	keine	<b>Sensitivität:</b> 82 % [63 %; 94 %]	452 [347; 517] <sup>f</sup>	534 [410; 611]	616 [473; 705]	sehr niedrig	
	FN										98 [33; 203] <sup>g</sup>		116 [39; 240]
	RN		nicht zu- treffend					schwer- wiegend <sup>h</sup>	Spezifität: 95 % [77 %; 100 %]	430 [347; 449] <sup>i</sup>	334 [270; 350]	239 [193; 250]	sehr niedrig
	FP										20 [1; 103] <sup>j</sup>	16 [0; 80]	
<b>≥ 3 klinische Parameter</b>													
1 (80) [40]	RP	sehr schwer- wiegend <sup>k</sup>	nicht zu- treffend	nicht schwer- wiegend	unent- deckt	schwer- wiegend <sup>l</sup>	keine	<b>Sensitivität:</b> 58 % [43 %; 71 %]	317 [238; 392]	375 [281; 463]	433 [324; 535]	sehr niedrig	
	FN										233 [158; 312]		275 [187; 369]
	RN		nicht zu- treffend					schwer- wiegend <sup>h</sup>	Spezifität: 79 % [59 %; 92 %]	354 [266; 413]	275 [207; 321]	196 [148; 229]	sehr niedrig
	FP										96 [37; 184]	75 [29; 143]	

Tabelle 9: Evidenzprofil zur diagnostischen Güte zu Dysphagie von klinischer Schluckuntersuchung im Vergleich zu instrumenteller Diagnostik (mehrsseitige Tabelle)

N (n)	Vierfelder- tafel-Ergeb- nisse	Faktoren der Qualität der Evidenz						Testgüte % [95 %-KI]	Schätzung für 1000 untersuchte Personen [95 %-KI]			Qualität der Evidenz <sup>a</sup>	
		Studien- limitationen <sup>b</sup>	Inkonsis- tenz	Indirekt- heit	Publika- tions- bias	Fehlende Genauig- keit	Andere Faktoren		Prävalenz- Annahme: 55 % <sup>c</sup>	Prävalenz- Annahme: 65 %	Prävalenz- Annahme: 75 %		
<b>≥ 2 klinische Parameter</b>													
1 (80) [40]	RP	sehr schwer- wiegend <sup>k</sup>	nicht zu- treffend	nicht schwer- wiegend	unent- deckt	nicht schwer- wiegend	keine	<b>Sensitivität:</b> 73 % [59 %; 84 %]	402 [324; 464]	475 [383; 549]	548 [442; 633]	niedrig	
	FN								148 [86; 226]	175 [101; 267]	202 [117; 308]		
	RN		nicht zu- treffend					schwer- wiegend <sup>m</sup>	<b>Spezifität:</b> 61 % [41 %; 78 %]	273 [183; 353]	213 [142; 275]	152 [101; 196]	sehr niedrig
	FP									177 [97; 267]	138 [75; 208]	98 [54; 149]	

Tabelle 9: Evidenzprofil zur diagnostischen Güte zu Dysphagie von klinischer Schluckuntersuchung im Vergleich zu instrumenteller Diagnostik (mehreseitige Tabelle)

N (n)	Vierfelder- tafel-Ergeb- nisse	Faktoren der Qualität der Evidenz						Testgüte % [95 %-KI]	Schätzung für 1000 untersuchte Personen [95 %-KI]			Qualität der Evidenz <sup>a</sup>
		Studien- limitationen <sup>b</sup>	Inkonsis- tenz	Indirekt- heit	Publika- tions- bias	Fehlende Genauig- keit	Andere Faktoren		Prävalenz- Annahme: 55 % <sup>c</sup>	Prävalenz- Annahme: 65 %	Prävalenz- Annahme: 75 %	
<p>a. Unter Qualität der Evidenz wird die studienübergreifende zielgrößenbezogene Qualität der Evidenz verstanden.</p> <p>b. Die Kriterien des Verzerrungspotenzials pro Studie sind Anhang A zu entnehmen.</p> <p>c. Die Festlegung der Prävalenz-Annahmen orientierte sich an den Prävalenzraten der eingeschlossenen Studien mit verwertbaren Daten. Exemplarische Erläuterung zur Interpretation: Eine Prävalenz-Annahme von 55 % bedeutet für 1000 untersuchte Personen, dass 550 Personen eine Dysphagie haben.</p> <p>d. Die Kriterien Indextest sowie Personenfluss und zeitlicher Ablauf wurden mit hoch bewertet. Zudem weisen die berichteten Ergebnisse folgende Diskrepanzen auf, welche in einer Autorenanfrage angesprochen, von den Autorinnen und Autoren jedoch nicht aufgelöst (Anhang B) wurden: zum einen entsprechen die in der Studie berichteten Werte zu Spezifität, positiv prädiktivem Wert und negativ prädiktivem Wert nicht den Werten, die auf Grundlage der Vierfeldertafel-Daten berechnet wurden. Zum anderen werden in der Studie unter den Ergebnissen zu der Zielgröße „yield“ mehr Patientinnen und Patienten mit Dysphagie angegeben als mit der VFSS laut Studie ermittelt wurden.</p> <p>e. Das 95 %-KI der Sensitivität überdeckt die Grenze von 90 %.</p> <p>f. Exemplarische Erläuterung zur Interpretation: Bei 347 bis 517 von 550 Personen wird richtigerweise eine Dysphagie diagnostiziert und die Personen erhalten eine instrumentelle Diagnostik.</p> <p>g. Exemplarische Erläuterung zur Interpretation: Bei 33 bis 203 von 550 Personen wird fälschlicherweise keine Dysphagie diagnostiziert und die Personen erhalten keine instrumentelle Diagnostik und keine Therapie.</p> <p>h. Das 95 %-KI der Spezifität überdeckt die Grenze von 80 %.</p> <p>i. Exemplarische Erläuterung zur Interpretation: Bei 347 bis 449 von 450 Personen wird richtigerweise diagnostiziert, dass sie keine Dysphagie haben und ihnen wird eine instrumentelle Diagnostik erspart.</p> <p>j. Exemplarische Erläuterung zur Interpretation: Bei 1 bis 103 von 450 Personen wird fälschlicherweise eine Dysphagie diagnostiziert und sie haben durch den Indextest einen weiteren Test zusätzlich zur instrumentellen Diagnostik.</p> <p>k. Das Kriterium Indextest wurde mit hoch bewertet.</p> <p>l. Das 95 %-KI der Sensitivität überdeckt die Grenze von 50 %.</p> <p>m. Das 95 %-KI der Spezifität überdeckt die Grenze von 50 %.</p> <p>FN: falsch-negativ (Personen, bei denen fälschlicherweise keine Dysphagie diagnostiziert wurde); FP: falsch-positiv (Personen, bei denen fälschlicherweise eine Dysphagie diagnostiziert wurde); KI: Konfidenzintervall; N: Anzahl der Studien; n: Anzahl ausgewerteter Personen; RN: richtig-negativ (Personen ohne Dysphagie); RP: richtig-positiv (Personen mit Dysphagie); VFSS: videofluoroskopische Schluckuntersuchung</p>												

### 5.3.2 Ergebnisse zur diagnostischen Güte: Referenzkriterium Aspiration

#### 5.3.2.1 Ergebnisse zum Wasserschlucktest im Vergleich zu instrumenteller Diagnostik

Tabelle 10: Evidenzprofil zur diagnostischen Güte zu Aspiration von Wasserschlucktest im Vergleich zu instrumenteller Diagnostik (mehrsseitige Tabelle)

N (n)	Vierfelder-tafel-Ergebnisse	Faktoren der Qualität der Evidenz						Testgüte % [95 %-KI]	Schätzung für 1000 untersuchte Personen [95 %-KI]			Qualität der Evidenz <sup>a</sup>
		Studien-limitationen <sup>b</sup>	Inkonsistenz	Indirektheit	Publikationsbias	Fehlende Genauigkeit	Andere Faktoren		Prävalenz-Annahme: 25 % <sup>c</sup>	Prävalenz-Annahme: 35 %	Prävalenz-Annahme: 45 %	
<b>51 ml WST</b>												
1 (184) [34]	RP	nicht schwerwiegend	nicht zutreffend	schwerwiegend <sup>d</sup>	unentdeckt	schwerwiegend <sup>e</sup>	keine	<b>Sensitivität:</b> 94 % [86 %; 98 %]	234 [215; 245] <sup>f</sup>	328 [301; 343]	422 [387; 441]	niedrig
	FN								16 [5; 35] <sup>g</sup>	22 [7; 49]	28 [9; 63]	
	RN	nicht zutreffend	schwerwiegend <sup>h</sup>	unentdeckt	schwerwiegend <sup>e</sup>	keine	<b>Spezifität:</b> 58 % [48 %; 67 %]	433 [357; 505] <sup>i</sup>	375 [309; 438]	317 [262; 370]	niedrig	
	FP							317 [245; 393] <sup>j</sup>	275 [212; 341]	233 [180; 288]		
<b>FraMaDySc</b>												
1 (184) [34]	RP	sehr schwerwiegend <sup>k</sup>	nicht zutreffend	schwerwiegend <sup>d</sup>	unentdeckt	schwerwiegend <sup>e</sup>	keine	<b>Sensitivität:</b> 94 % [86 %; 98 %]	234 [215; 245]	328 [301; 343]	422 [387; 441]	sehr niedrig
	FN								16 [5; 35]	22 [7; 49]	28 [9; 63]	
	RN	nicht zutreffend	schwerwiegend <sup>h</sup>	unentdeckt	schwerwiegend <sup>e</sup>	keine	<b>Spezifität:</b> 60 % [49,5 %; 69 %]	447 [372; 518]	388 [322; 449]	328 [272; 380]	sehr niedrig	
	FP							303 [232; 378]	263 [201; 328]	222 [170; 278]		

Tabelle 10: Evidenzprofil zur diagnostischen Güte zu Aspiration von Wasserschlucktest im Vergleich zu instrumenteller Diagnostik (mehrsseitige Tabelle)

N (n)	Vierfelder- tafel- Ergebnisse	Faktoren der Qualität der Evidenz						Testgüte % [95 %-KI]	Schätzung für 1000 untersuchte Personen [95 %-KI]			Qualität der Evidenz <sup>a</sup>
		Studien- limitationen <sup>b</sup>	Inkonsis- tenz	Indirekt- heit	Publika- tionsbias	Fehlende Genauigkeit	Andere Faktoren		Prävalenz- Annahme: 25 % <sup>c</sup>	Prävalenz- Annahme: 35 %	Prävalenz- Annahme: 45 %	
<b>modifizierter WST</b>												
1 (28) [36]	RP	sehr schwer- wiegend <sup>l</sup>	nicht zu- treffend	schwer- wiegend <sup>d</sup>	unentdeckt	sehr schwer- wiegend <sup>m</sup>	keine	<b>Sensitivität:</b> 90 % [55 %; 100 %]	225 [139; 249]	315 [194; 349]	405 [250; 449]	sehr niedrig
	FN								25 [1; 111]	35 [1; 156]	45 [1; 200]	
	RN	nicht zu- treffend	sehr schwer- wiegend <sup>n</sup>	keine	<b>Spezifität:</b> 61 % [36 %; 83 %]	458 [268; 620]	397 [232; 538]	336 [197; 455]	sehr niedrig			
	FP					292 [130; 482]	253 [112; 418]	214 [95; 353]				
<b>100 ml WST</b>												
1 (155) [38]	RP	schwer- wiegend <sup>o</sup>	nicht zu- treffend	sehr schwer- wiegend <sup>p</sup>	unentdeckt	nicht schwer- wiegend	keine	<b>Sensitivität:</b> 77 % [60 %; 89,6 %]	193 [150; 224]	270 [210; 314]	347 [269; 403]	sehr niedrig
	FN								57 [26; 100]	80 [36; 140]	103 [47; 181]	
	RN	nicht zu- treffend	schwer- wiegend <sup>h</sup>	keine	<b>Spezifität:</b> 47 % [38 %; 56 %]	350 [281; 420]	303 [244; 364]	257 [206; 308]	sehr niedrig			
	FP					400 [330; 469]	347 [286; 406]	293 [242; 344]				

Tabelle 10: Evidenzprofil zur diagnostischen Güte zu Aspiration von Wasserschlucktest im Vergleich zu instrumenteller Diagnostik (mehrsseitige Tabelle)

N (n)	Vierfelder- tafel- Ergebnisse	Faktoren der Qualität der Evidenz						Testgüte % [95 %-KI]	Schätzung für 1000 untersuchte Personen [95 %-KI]			Qualität der Evidenz <sup>a</sup>
		Studien- limitationen <sup>b</sup>	Inkonsis- tenz	Indirekt- heit	Publika- tionsbias	Fehlende Genauigkeit	Andere Faktoren		Prävalenz- Annahme: 25 % <sup>c</sup>	Prävalenz- Annahme: 35 %	Prävalenz- Annahme: 45 %	
<p>a. Unter Qualität der Evidenz wird die studienübergreifende zielgrößenbezogene Qualität der Evidenz verstanden.</p> <p>b. Die Kriterien des Verzerrungspotenzials pro Studie sind Anhang A zu entnehmen.</p> <p>c. Die Festlegung der Prävalenz-Annahmen orientierte sich an den Prävalenzraten der eingeschlossenen Studien mit verwertbaren Daten. Exemplarische Erläuterung zur Interpretation: Eine Prävalenz-Annahme von 25 % bedeutet für 1000 untersuchte Personen, dass 250 Personen eine Aspiration haben.</p> <p>d. Das Referenzkriterium Aspiration untersucht nur einen Teilaspekt der Dysphagie.</p> <p>e. Das 95 %-KI der Sensitivität überdeckt die Grenze von 90 %.</p> <p>f. Exemplarische Erläuterung zur Interpretation: Bei 215 bis 245 von 250 Personen wird richtigerweise eine Aspiration diagnostiziert und die Personen erhalten eine instrumentelle Diagnostik.</p> <p>g. Exemplarische Erläuterung zur Interpretation: Bei 5 bis 35 von 250 Personen wird fälschlicherweise keine Aspiration diagnostiziert und die Personen erhalten keine instrumentelle Diagnostik und keine Therapie.</p> <p>h. Das 95 %-KI der Spezifität überdeckt die Grenze von 50 %.</p> <p>i. Exemplarische Erläuterung zur Interpretation: Bei 357 bis 505 von 750 Personen wird richtigerweise diagnostiziert, dass sie keine Aspiration haben und ihnen wird eine instrumentelle Diagnostik erspart.</p> <p>j. Exemplarische Erläuterung zur Interpretation: Bei 245 bis 393 von 750 Personen wird fälschlicherweise eine Aspiration diagnostiziert und sie haben durch den Indextest einen weiteren Test zusätzlich zur instrumentellen Diagnostik.</p> <p>k. Das Kriterium Indextest wurde hoch bewertet.</p> <p>l. Die Kriterien Personenselektion, Indextest sowie Personenfluss und zeitlicher Ablauf wurden mit hoch bewertet.</p> <p>m. Das 95 %-KI der Sensitivität überdeckt die 90 % und die Studie ist sehr klein.</p> <p>n. Das 95 %-KI der Spezifität überdeckt die Grenzen von 50 % und 80 %.</p> <p>o. Die Kriterien Referenzstandard sowie Personenfluss und zeitlicher Ablauf wurden mit hoch bewertet.</p> <p>p. Das Referenzkriterium Aspiration untersucht nur einen Teilaspekt der Dysphagie. Das Referenzkriterium Aspiration wurde in der Publikation nicht definiert, sodass es nicht repliziert werden kann.</p> <p>FN: falsch-negativ (Personen, bei denen fälschlicherweise keine Aspiration diagnostiziert wurde); FP: falsch-positiv (Personen, bei denen fälschlicherweise eine Aspiration diagnostiziert wurde); FraMaDySc: Frankfurt/Marburg Dysphagia Screening; KI: Konfidenzintervall; N: Anzahl der Studien; n: Anzahl ausgewerteter Personen; RN: richtig-negativ (Personen ohne Aspiration); RP: richtig-positiv (Personen mit Aspiration); WST: Wasserschlucktest</p>												

### 5.3.2.2 Ergebnisse zur klinischen Schluckuntersuchung im Vergleich zu instrumenteller Diagnostik

Tabelle 11: Evidenzprofil zur diagnostischen Güte zu Aspiration von klinischer Schluckuntersuchung im Vergleich zu instrumenteller Diagnostik (mehrsseitige Tabelle)

N (n)	Vierfelder- tafel-Ergeb- nisse	Faktoren der Qualität der Evidenz						Testgüte % [95 %-KI]	Schätzung für 1000 untersuchte Personen [95 %-KI]			Qualität der Evidenz <sup>a</sup>			
		Studien- limitationen <sup>b</sup>	Inkonsis- tenz	Indirekt- heit	Publikations- bias	Fehlende Genauigkeit	Andere Faktoren		Prävalenz- Annahme: 25 % <sup>c</sup>	Prävalenz- Annahme: 35 %	Prävalenz- Annahme: 45 %				
<b>≥ 3 klinische Parameter</b>															
1 (80) [40]	RP	sehr schwer- wiegend <sup>d</sup>	nicht zu- treffend	schwer- wiegend <sup>e</sup>	unentdeckt	schwer- wiegend <sup>f</sup>	keine	<b>Sensitivität:</b> 59 % [42 %; 74 %]	147 [105; 186] <sup>g</sup>	206 [147; 261]	265 [189; 335]	sehr niedrig			
	FN												103 [64; 145] <sup>h</sup>	144 [89; 203]	185 [115; 261]
	RN		nicht zu- treffend								<b>Spezifität:</b> 68 % [52 %; 82 %]	512 [389; 614] <sup>j</sup>	444 [337; 532]	376 [286; 451]	sehr niedrig
	FP										238 [136; 361] <sup>k</sup>	206 [118; 313]	174 [99; 264]		
<b>≥ 2 klinische Parameter</b>															
1 (80) [40]	RP	sehr schwer- wiegend <sup>d</sup>	nicht zu- treffend	schwer- wiegend <sup>e</sup>	unentdeckt	nicht schwer- wiegend	keine	<b>Sensitivität:</b> 77 % [61 %; 89 %]	192 [152; 222]	269 [212; 311]	346 [273; 400]	sehr niedrig			
	FN												58 [28; 98]	81 [39; 138]	104 [50; 177]
	RN		nicht zu- treffend								<b>Spezifität:</b> 54 % [37 %; 69 %]	402 [281; 520]	349 [243; 451]	295 [206; 381]	sehr niedrig
	FP										348 [230; 469]	301 [199; 407]	255 [169; 344]		

Tabelle 11: Evidenzprofil zur diagnostischen Güte zu Aspiration von klinischer Schluckuntersuchung im Vergleich zu instrumenteller Diagnostik (mehreseitige Tabelle)

N (n)	Vierfelder- tafel-Ergeb- nisse	Faktoren der Qualität der Evidenz						Testgüte % [95 %-KI]	Schätzung für 1000 untersuchte Personen [95 %-KI]			Qualität der Evidenz <sup>a</sup>
		Studien- limitationen <sup>b</sup>	Inkonsis- tenz	Indirekt- heit	Publikations- bias	Fehlende Genauigkeit	Andere Faktoren		Prävalenz- Annahme: 25 % <sup>c</sup>	Prävalenz- Annahme: 35 %	Prävalenz- Annahme: 45 %	
<p>a. Unter Qualität der Evidenz wird die studienübergreifende zielgrößenbezogene Qualität der Evidenz verstanden.</p> <p>b. Die Kriterien des Verzerrungspotenzials pro Studie pro Zielgröße sind Anhang A zu entnehmen.</p> <p>c. Die Festlegung der Prävalenz-Annahmen orientierte sich an den Prävalenzraten der eingeschlossenen Studien mit verwertbaren Daten. Exemplarische Erläuterung zur Interpretation: Eine Prävalenz-Annahme von 25 % bedeutet für 1000 untersuchte Personen, dass 250 Personen eine Aspiration haben.</p> <p>d. Das Kriterium Indextest wurde mit hoch bewertet.</p> <p>e. Das Referenzkriterium Aspiration untersucht nur einen Teilaspekt der Dysphagie.</p> <p>f. Das 95 %-KI der Sensitivität überdeckt die Grenze von 50 %.</p> <p>g. Exemplarische Erläuterung zur Interpretation: Bei 105 bis 186 von 250 Personen wird richtigerweise eine Aspiration diagnostiziert und die Personen erhalten eine instrumentelle Diagnostik.</p> <p>h. Exemplarische Erläuterung zur Interpretation: Bei 64 bis 145 von 250 Personen wird fälschlicherweise keine Aspiration diagnostiziert und die Personen erhalten keine instrumentelle Diagnostik und keine Therapie.</p> <p>i. Das 95 %-KI der Spezifität überdeckt die Grenze von 80 %.</p> <p>j. Exemplarische Erläuterung zur Interpretation: Bei 389 bis 614 von 750 Personen wird richtigerweise diagnostiziert, dass sie keine Aspiration haben und ihnen wird eine instrumentelle Diagnostik erspart.</p> <p>k. Exemplarische Erläuterung zur Interpretation: Bei 136 bis 361 von 750 Personen wird fälschlicherweise eine Aspiration diagnostiziert und sie haben durch den Indextest einen weiteren Test zusätzlich zur instrumentellen Diagnostik.</p> <p>l. Das 95 %-KI der Spezifität überdeckt die Grenze von 50 %.</p> <p>FN: falsch-negativ (Personen, bei denen fälschlicherweise keine Aspiration diagnostiziert wurde); FP: falsch-positiv (Personen, bei denen fälschlicherweise eine Aspiration diagnostiziert wurde); KI: Konfidenzintervall; N: Anzahl der Studien; n: Anzahl ausgewerteter Personen; RN: richtig-negativ (Personen ohne Aspiration); RP: richtig-positiv (Personen mit Aspiration)</p>												

### 5.3.3 Ergebnisse zur diagnostischen Güte: Referenzkriterium Oralisierungseinschränkung

#### 5.3.3.1 Ergebnisse zum Wasserschlucktest im Vergleich zu instrumenteller Diagnostik

Tabelle 12: Evidenzprofil zur diagnostischen Güte zu Oralisierungseinschränkung von Wasserschlucktest im Vergleich zu instrumenteller Diagnostik (mehreseitige Tabelle)

N (n)	Vierfelder-tafel-Ergebnisse	Faktoren der Qualität der Evidenz						Testgüte % [95 %-KI]	Schätzung für 1000 untersuchte Personen [95 %-KI]			Qualität der Evidenz <sup>a</sup>	
		Studien-limitationen <sup>b</sup>	Inkonsistenz	Indirektheit	Publikations-bias	Fehlende Genauigkeit	Andere Faktoren		Prävalenz-Annahme: 45 % <sup>c</sup>	Prävalenz-Annahme: 55 %	Prävalenz-Annahme: 65 %		
<b>51 ml WST</b>													
1 (184) [34]	RP	nicht schwerwiegend	nicht zutreffend	schwerwiegend <sup>d</sup>	unentdeckt	schwerwiegend <sup>e</sup>	keine	<b>Sensitivität:</b> 91 % [84 %; 96 %]	411 [378; 432] <sup>f</sup>	502 [462; 528]	593 [546; 624]	niedrig	
	FN								39 [18; 72] <sup>g</sup>	48 [22; 88]	57 [26; 104]		
	RN		nicht zutreffend					nicht schwerwiegend	<b>Spezifität:</b> 69 % [58 %; 79 %]	380 [318; 434] <sup>h</sup>	311 [261; 355]	242 [203; 276]	moderat
	FP									170 [116; 232] <sup>i</sup>	139 [95; 189]	108 [74; 147]	
<b>FraMaDySc</b>													
1 (184) [34]	RP	sehr schwerwiegend <sup>j</sup>	nicht zutreffend	schwerwiegend <sup>d</sup>	unentdeckt	schwerwiegend <sup>e</sup>	keine	<b>Sensitivität:</b> 91 % [84 %; 96 %]	411 [378; 432]	502 [462; 528]	593 [546; 624]	sehr niedrig	
	FN								39 [18; 72]	48 [22; 88]	57 [26; 104]		
	RN		nicht zutreffend					schwerwiegend <sup>k</sup>	<b>Spezifität:</b> 72 % [60 %; 81 %]	394 [333; 446]	322 [272; 365]	251 [212; 284]	sehr niedrig
	FP									156 [104; 217]	128 [85; 178]	99 [66; 138]	

Tabelle 12: Evidenzprofil zur diagnostischen Güte zu Oralisierungseinschränkung von Wasserschlucktest im Vergleich zu instrumenteller Diagnostik (mehrseitige Tabelle)

N (n)	Vierfelder- tafel- Ergebnisse	Faktoren der Qualität der Evidenz					Testgüte % [95 %-KI]	Schätzung für 1000 untersuchte Personen [95 %-KI]			Qualität der Evidenz <sup>a</sup>
		Studien- limitationen <sup>b</sup>	Inkonsis- tenz	Indirekt- heit	Publikations- bias	Fehlende Genauigkeit		Andere Faktoren	Prävalenz- Annahme: 45 % <sup>c</sup>	Prävalenz- Annahme: 55 %	
<p>a. Unter Qualität der Evidenz wird die studienübergreifende zielgrößenbezogene Qualität der Evidenz verstanden.</p> <p>b. Die Kriterien des Verzerrungspotenzials pro Studie sind Anhang A zu entnehmen.</p> <p>c. Die Festlegung der Prävalenz-Annahmen orientierte sich an den Prävalenzraten der eingeschlossenen Studien mit verwertbaren Daten. Exemplarische Erläuterung zur Interpretation: Eine Prävalenz-Annahme von 45 % bedeutet für 1000 untersuchte Personen, dass 450 Personen eine Oralisierungseinschränkung haben.</p> <p>d. Das Referenzkriterium Oralisierungseinschränkung untersucht nur einen Teilaspekt der Dysphagie.</p> <p>e. Das 95 %-KI der Sensitivität überdeckt die Grenze von 90 %.</p> <p>f. Exemplarische Erläuterung zur Interpretation: Bei 378 bis 432 von 450 Personen wird richtigerweise eine Oralisierungseinschränkung diagnostiziert und die Personen erhalten eine instrumentelle Diagnostik.</p> <p>g. Exemplarische Erläuterung zur Interpretation: Bei 18 bis 72 von 450 Personen wird fälschlicherweise keine Oralisierungseinschränkung diagnostiziert und die Personen erhalten keine instrumentelle Diagnostik und keine Therapie.</p> <p>h. Exemplarische Erläuterung zur Interpretation: Bei 318 bis 434 von 550 Personen wird richtigerweise diagnostiziert, dass sie keine Oralisierungseinschränkung haben und ihnen wird eine instrumentelle Diagnostik erspart.</p> <p>i. Exemplarische Erläuterung zur Interpretation: Bei 116 bis 232 von 550 Personen wird fälschlicherweise eine Oralisierungseinschränkung diagnostiziert und sie haben durch den Indextest einen weiteren Test zusätzlich zur instrumentellen Diagnostik.</p> <p>j. Das Kriterium Indextest wurde mit hoch bewertet.</p> <p>k. Das 95 %-KI der Spezifität überdeckt die Grenze von 80 %.</p> <p>FN: falsch-negativ (Personen, bei denen fälschlicherweise keine Oralisierungseinschränkung diagnostiziert wurde); FP: falsch-positiv (Personen, bei denen fälschlicherweise eine Oralisierungseinschränkung diagnostiziert wurde); FraMaDySc: Frankfurt/Marburg Dysphagia Screening; KI: Konfidenzintervall; N: Anzahl der Studien; n: Anzahl ausgewerteter Personen; RN: richtig-negativ (Personen ohne Oralisierungseinschränkung); RP: richtig-positiv (Personen mit Oralisierungseinschränkung); WST: Wasserschlucktest</p>											

### 5.3.3.2 Ergebnisse zur klinischen Schluckuntersuchung im Vergleich zu instrumenteller Diagnostik

Tabelle 13: Evidenzprofil zur diagnostischen Güte zu Oralisierungseinschränkung von klinischer Schluckuntersuchung im Vergleich zu instrumenteller Diagnostik (mehrseitige Tabelle)

N (n)	Vierfelder- tafel-Ergeb- nisse	Faktoren der Qualität der Evidenz						Testgüte % [95 %-KI]	Schätzung für 1000 untersuchte Personen [95 %-KI]			Qualität der Evidenz <sup>a</sup>
		Studien- limitationen <sup>b</sup>	Inkonsis- tenz	Indirekt- heit	Publikations- bias	Fehlende Genauig- keit	Andere Faktoren		Prävalenz- Annahme: 45 % <sup>c</sup>	Prävalenz- Annahme: 55 %	Prävalenz- Annahme: 65 %	
<b>≥ 3 klinische Parameter</b>												
1 (80) [40]	RP	sehr schwer- wiegend <sup>d</sup>	nicht zu- treffend	schwer- wiegend <sup>e</sup>	unentdeckt	schwer- wiegend <sup>f</sup>	keine	<b>Sensitivität:</b> 60 % [44 %; 74 %]	270 [199; 334] <sup>g</sup>	330 [244; 409]	390 [288; 483]	sehr niedrig
	FN								180 [116; 251] <sup>h</sup>	220 [141; 306]	260 [167; 362]	
	RN	nicht zu- treffend	schwer- wiegend <sup>i</sup>	keine	<b>Spezifität:</b> 74 % [57 %; 88 %]	409 [312; 481] <sup>j</sup>	334 [255; 394]	260 [199; 306]	sehr niedrig			
	FP					141 [69; 238] <sup>k</sup>	116 [56; 195]	90 [44; 151]				
<b>≥ 2 klinische Parameter</b>												
1 (80) [40]	RP	sehr schwer- wiegend <sup>d</sup>	nicht zu- treffend	schwer- wiegend <sup>e</sup>	unentdeckt	nicht schwer- wiegend	keine	<b>Sensitivität:</b> 78 % [63 %; 89 %]	350 [283; 400]	428 [346; 488]	506 [409; 577]	sehr niedrig
	FN								100 [50; 167]	122 [62; 204]	144 [73; 241]	
	RN	nicht zu- treffend	schwer- wiegend <sup>l</sup>	keine	<b>Spezifität:</b> 60 % [42 %; 76 %]	330 [232; 419]	270 [190; 343]	210 [147; 266]	sehr niedrig			
	FP					220 [131; 318]	180 [107; 260]	140 [84; 203]				

Tabelle 13: Evidenzprofil zur diagnostischen Güte zu Oralisierungseinschränkung von klinischer Schluckuntersuchung im Vergleich zu instrumenteller Diagnostik (mehrseitige Tabelle)

N (n)	Vierfelder- tafel-Ergeb- nisse	Faktoren der Qualität der Evidenz						Testgüte % [95 %-KI]	Schätzung für 1000 untersuchte Personen [95 %-KI]			Qualität der Evidenz <sup>a</sup>
		Studien- limitationen <sup>b</sup>	Inkonsis- tenz	Indirekt- heit	Publikations- bias	Fehlende Genauig- keit	Andere Faktoren		Prävalenz- Annahme: 45 % <sup>c</sup>	Prävalenz- Annahme: 55 %	Prävalenz- Annahme: 65 %	
<p>a. Unter Qualität der Evidenz wird die studienübergreifende zielgrößenbezogene Qualität der Evidenz verstanden.</p> <p>b. Die Kriterien des Verzerrungspotenzials pro Studie sind Anhang A zu entnehmen.</p> <p>c. Die Festlegung der Prävalenz-Annahmen orientierte sich an den Prävalenzraten der eingeschlossenen Studien mit verwertbaren Daten. Exemplarische Erläuterung zur Interpretation: Eine Prävalenz-Annahme von 45 % bedeutet für 1000 untersuchte Personen, dass 450 Personen eine Oralisierungseinschränkung haben.</p> <p>d. Das Kriterium Indextest wurde mit hoch bewertet.</p> <p>e. Das Referenzkriterium Oralisierungseinschränkung untersucht nur einen Teilaspekt der Dysphagie.</p> <p>f. Das 95 %-KI der Sensitivität überdeckt die Grenze von 50 %.</p> <p>g. Exemplarische Erläuterung zur Interpretation: Bei 199 bis 334 von 450 Personen wird richtigerweise eine Oralisierungseinschränkung diagnostiziert und die Personen erhalten eine instrumentelle Diagnostik.</p> <p>h. Exemplarische Erläuterung zur Interpretation: Bei 116 bis 251 von 450 Personen wird fälschlicherweise keine Oralisierungseinschränkung diagnostiziert und die Personen erhalten keine instrumentelle Diagnostik und keine Therapie.</p> <p>i. Das 95 %-KI der Spezifität überdeckt die Grenze von 80 %.</p> <p>j. Exemplarische Erläuterung zur Interpretation: Bei 312 bis 481 von 550 Personen wird richtigerweise diagnostiziert, dass sie keine Oralisierungseinschränkung haben und ihnen wird eine instrumentelle Diagnostik erspart.</p> <p>k. Exemplarische Erläuterung zur Interpretation: Bei 69 bis 238 von 550 Personen wird fälschlicherweise eine Oralisierungseinschränkung diagnostiziert und sie haben durch den Indextest einen weiteren Test zusätzlich zur instrumentellen Diagnostik.</p> <p>l. Das 95 %-KI der Spezifität überdeckt die Grenze von 50 %.</p> <p>FN: falsch-negativ (Personen, bei denen fälschlicherweise keine Oralisierungseinschränkung diagnostiziert wurde); FP: falsch-positiv (Personen, bei denen fälschlicherweise eine Oralisierungseinschränkung diagnostiziert wurde); KI: Konfidenzintervall; N: Anzahl der Studien; n: Anzahl ausgewerteter Personen; RN: richtig-negativ (Personen ohne Oralisierungseinschränkung); RP: richtig-positiv (Personen mit Oralisierungseinschränkung)</p>												

### 5.3.4 Ergebnisse zum negativ prädiktiven Wert

Der negativ prädiktive Wert wurde pro Referenzkriterium für die 3 Prävalenzen berechnet, welche bereits in den Evidenzprofilen verwendet wurden. Die Ergebnisse sind in Tabelle 14, Tabelle 15 und Tabelle 16 dargestellt.

Tabelle 14: Übersicht der negativ prädiktiven Werte zum Referenzkriterium Dysphagie

Vergleich Indextest N (n)	NPV % [95 %-KI]: (Anteil der Personen mit negativem Testergebnis, die tatsächlich keine Dysphagie haben)		
	Prävalenz-Annahme: 55 % (550 von 1000 untersuchte Personen mit Dysphagie)	Prävalenz-Annahme: 65 % (650 von 1000 untersuchte Personen mit Dysphagie)	Prävalenz-Annahme: 75 % (750 von 1000 untersuchte Personen mit Dysphagie)
<b>Wasserschlucktest im Vergleich zu instrumenteller Diagnostik</b>			
<b>51 ml WST</b>			
1 (184) [34]	88 % [81 %; 93 %]	83 % [74 %; 90 %]	75 % [63 %; 84 %]
<b>FraMaDySc</b>			
1 (184) [34]	89 % [82 %; 93 %]	84 % [74 %; 90 %]	76 % [64 %; 85 %]
<b>klinische Schluckuntersuchung im Vergleich zu instrumenteller Diagnostik</b>			
<b>MASA-C</b>			
1 (50) [39]	81 % [66 %; 91 %]	74 % [56 %; 86 %]	64 % [44 %; 80 %]
<b>≥ 3 klinische Parameter</b>			
1 (80) [40]	60 % [51 %; 69 %]	50 % [41 %; 59 %]	38 % [30 %; 47 %]
<b>≥ 2 klinische Parameter</b>			
1 (80) [40]	65 % [52 %; 76 %]	55 % [41 %; 68 %]	43 % [31 %; 56 %]
FraMaDySc: Frankfurt/Marburg Dysphagie Screening; KI: Konfidenzintervall; MASA-C: Mann Assessment of Swallowing Ability-Cancer; N: Anzahl der Studien; n: Anzahl ausgewerteter Personen; NPV: negativ prädiktiver Wert; WST: Wasserschlucktest			

Tabelle 15: Übersicht der negativ prädiktiven Werte zum Referenzkriterium Aspiration

Vergleich Indextest N (n)	NPV % [95 %-KI] (Anteil der Personen mit negativem Testergebnis, die tatsächlich keine Aspiration haben)		
	Prävalenz-Annahme: 25 % (250 von 1000 untersuchten Personen mit Aspiration)	Prävalenz-Annahme: 35 % (350 von 1000 untersuchten Personen mit Aspiration)	Prävalenz-Annahme: 45 % (450 von 1000 untersuchten Personen mit Aspiration)
<b>Wasserschlucktest im Vergleich zu instrumenteller Diagnostik</b>			
<b>51 ml WST</b>			
1 (184) [34]	97 % [92 %; 99 %]	94 % [88 %; 98 %]	92 % [83 %; 96 %]
<b>FraMaDySc</b>			
1 (184) [34]	97 % [92 %; 99 %]	95 % [88 %; 98 %]	92 % [83 %; 97 %]
<b>modifizierter WST</b>			
1 (28) [36]	95 % [73 %; 99 %]	92 % [63 %; 99 %]	88 % [53 %; 98 %]
<b>100 ml WST</b>			
1 (155) [38]	86 % [76 %; 92 %]	79 % [67 %; 88 %]	71 % [57 %; 83 %]
<b>klinische Schluckuntersuchung im Vergleich zu instrumenteller Diagnostik</b>			
<b>≥ 3 klinische Parameter</b>			
1 (80) [40]	83 % [76 %; 88 %]	76 % [67 %; 83 %]	67 % [57 %; 76 %]
<b>≥ 2 klinische Parameter</b>			
1 (80) [40]	87 % [79 %; 93 %]	81 % [69 %; 89 %]	74 % [60 %; 84 %]
FraMaDySc: Frankfurt/Marburg Dysphagie Screening; KI: Konfidenzintervall; N: Anzahl der Studien; n: Anzahl ausgewerteter Personen; NPV: negativ prädiktiver Wert; WST: Wasserschlucktest			

Tabelle 16: Übersicht der negativ prädiktiven Werte zum Referenzkriterium Oralisierungseinschränkung

Vergleich Indextest N (n)	NPV % [95 %-KI] (Anteil der Personen mit negativem Testergebnis, die tatsächlich keine Oralisierungseinschränkung haben)		
	Prävalenz-Annahme: 45 % (450 von 1000 untersuchten Personen mit Oralisierungseinschränkung)	Prävalenz-Annahme: 55 % (550 von 1000 untersuchten Personen mit Oralisierungseinschränkung)	Prävalenz-Annahme: 65 % (650 von 1000 untersuchten Personen mit Oralisierungseinschränkung)
<b>Wasserschlucktest im Vergleich zu instrumenteller Diagnostik</b>			
<b>51 ml WST</b>			
1 (184) [34]	91 % [84 %; 95 %]	87 % [77 %; 92 %]	81 % [69 %; 89 %]
<b>FraMaDySc</b>			
1 (184) [34]	91 % [84 %; 95 %]	87 % [78 %; 93 %]	82 % [70 %; 89 %]
<b>klinische Schluckuntersuchung im Vergleich zu instrumenteller Diagnostik</b>			
<b>≥ 3 klinische Parameter</b>			
1 (80) [40]	69 % [60 %; 77 %]	60 % [50 %; 70 %]	50 % [40 %; 60 %]
<b>≥ 2 klinische Parameter</b>			
1 (80) [40]	77 % [64 %; 86 %]	69 % [55 %; 80 %]	59 % [44 %; 73 %]
FraMaDySc: Frankfurt/Marburg Dysphagie Screening; KI: Konfidenzintervall; N: Anzahl der Studien; n: Anzahl ausgewerteter Personen; NPV: negativ prädiktiver Wert; WST: Wasserschlucktest			

## **6 Zusammenfassung und Anmerkungen zu den Ergebnissen**

### **6.1 Zusammenfassung der Ergebnisse**

Dieser Evidenzbericht bezieht sich auf Patientinnen und Patienten ( $\geq 18$  Jahre) mit Kopf-Hals-Tumor während / nach Strahlen-(Chemo-) Therapie oder nach Tumorresektion. Für die Fragestellung zu dieser Population wurde der Einsatz von 3 verschiedenen Triage-Tests vor Einsatz instrumenteller Diagnostik bei Testpositiven im Vergleich zum alleinigen Einsatz instrumenteller Diagnostik zur Indikation für eine Therapie der Dysphagie untersucht. Zu dieser Fragestellung wurden keine Studien zur diagnostisch-therapeutischen Kette identifiziert. Für den Baustein der Fragestellung zur diagnostischen Güte wurden jeweils 3 Studien identifiziert, welche Evidenz von Wasserschlucktest bzw. klinischer Schluckuntersuchung im Vergleich zur instrumentellen Diagnostik berichteten. Die Daten aus 1 Studie zur klinischen Schluckuntersuchung waren nicht verwertbar. Es wurden keine Studien zur diagnostischen Güte für den Indextest Fragebogen zur subjektiven Einschätzung des Schluckvermögens identifiziert.

#### **Wasserschlucktest im Vergleich zu instrumenteller Diagnostik**

In 3 Studien wurde die diagnostische Güte von 4 unterschiedlichen Wasserschlucktests untersucht. Die in den Studien herangezogenen Referenztests unterschieden sich bezüglich der getesteten Kriterien Dysphagie, Aspiration und / oder Oralisierungseinschränkung.

Für 2 Indextests lag die diagnostische Güte zum Referenzkriterium Dysphagie vor. Obwohl das Frankfurt/Marburg Dysphagie Screening (FraMaDySc) die höchste diagnostische Güte erreichte war die Qualität der Evidenz niedrig. Zudem umfassten die 95 %-KIs der Sensitivität und Spezifität die Empfehlungsgrenzen von 90 % bzw. 80 %. Der 51 ml Wasserschlucktest (WST) wies zwar dieselbe Sensitivität wie der FraMaDySc auf, die Spezifität war jedoch geringer (moderate Qualität der Evidenz).

Für das Referenzkriterium Oralisierungseinschränkung lag die diagnostische Güte ebenfalls für 2 Indextests vor. Sowohl für den 51 ml WST als auch für den FraMaDySc war die Sensitivität vergleichbar mit der für das Referenzkriterium Dysphagie (niedrige bzw. sehr niedrige Qualität der Evidenz). Das 95 %-KI der Spezifität war für beide Tests niedriger im Vergleich zum Referenzkriterium Dysphagie (moderate bzw. sehr niedrige Qualität der Evidenz).

Das Referenzkriterium Aspiration wurde mit 4 Indextests untersucht. Die 95 %-KIs der Spezifität aller Indextests umfassten jeweils die 50 %-Grenze, sodass unklar ist, ob sie als Triage-Test für die Erkennung von Personen, die keine Aspiration aufweisen, als brauchbar für die klinische Praxis eingeschätzt werden können (niedrige bis sehr niedrige Qualität der Evidenz). Der 51 ml WST und der FraMaDySc zeigten die höchste Sensitivität, jedoch umfassten auch hier die 95 %-KIs die Empfehlungsgrenze von 90 % (niedrige bzw. sehr niedrige Qualität der Evidenz). Das 95 %-KI der Sensitivität lag für den 100 ml WST vollständig

unterhalb der Empfehlungsgrenze wohingegen das des modifizierten WST sehr ungenau war und sowohl die Empfehlungsgrenze umfasste als auch lediglich knapp über der 50 %-Grenze lag (sehr niedrige Qualität der Evidenz).

Ergänzend zu den prävalenzunabhängigen Ergebnissen wurden die negativen prädiktiven Werte (NPVs) dargestellt. Tendenziell sank der NPV mit steigender Prävalenzrate. Bei einer Prävalenz-Annahme von 25 % für das Referenzkriterium Aspiration lagen die 95 %-KIs des 51 ml WST und des FraMaDySc über 90 %. Bei dieser Prävalenz-Annahme kann davon ausgegangen werden, dass weniger als 10 % der Personen mit negativem Testergebnis ein falsch-negatives Testergebnis (also eine Aspiration) haben. Zu beachten ist, dass die Referenzkriterien Dysphagie und Oralisierungseinschränkung niedrigere NPVs aufwiesen. Bei den unterschiedlichen Wasserschlucktests fällt auf, dass die NPVs zum modifizierten WST und 100 ml WST lagen unter denen zum 51 ml WST und FraMaDySc.

### **Klinische Schluckuntersuchung im Vergleich zu instrumenteller Diagnostik**

Aus 2 Studien wurden Daten zur diagnostischen Güte von 3 klinischen Schluckuntersuchungen herangezogen. Auch für diese Indextests wurden Ergebnisse zu den 3 Referenzkriterien Dysphagie, Aspiration und / oder Oralisierungseinschränkung berichtet. Die Qualität der Evidenz für alle Ergebnisse war bis auf eine Ausnahme sehr niedrig. Die höchste diagnostische Güte erzielte der MASA-C für das Referenzkriterium Dysphagie. Sowohl für Spezifität als auch für Sensitivität umfassten die 95 %-KIs die Empfehlungsgrenzen. Das 95 %-KI des NPV des MASA-C für eine Prävalenz-Annahme von 55 % lag zwischen 66 % und 91 %. Für die beiden anderen Indextests  $\geq 3$  und  $\geq 2$  klinische Parameter waren die Sensitivität und die Spezifität, unabhängig vom Referenzkriterium, niedriger als für den MASA-C, wobei jeweils eines der 95 %-KI zu Sensitivität oder Spezifität die 50 % Grenze umfasste und das 95 %-KI meist vollständig unter der jeweiligen Empfehlungsgrenze lag. Auch die NPV lagen für die dargestellten Prävalenz-Annahmen unter denen des MASA-C.

### **6.2 Anmerkungen zu den Ergebnissen**

Unabhängig vom Referenzkriterium lagen bei allen Indextests die beiden 95 %-KI-Grenzen weder für Sensitivität noch für Spezifität über der jeweiligen festgelegten Empfehlungsgrenze von 90 % bzw. 80 %. Nach diesem Maßstab könnte keiner der Indextest für die klinische Praxis empfohlen werden (je nach Indextest moderate bis sehr niedrige Qualität der Evidenz). Ergebnisse zum NPV wurden ergänzend zu den Prävalenz-unabhängigen Maßen Sensitivität und Spezifität dargestellt, weshalb für den NPV keine Bewertung der Qualität der Evidenz vorgenommen wurde.

Dieser Evidenzbericht stellt neben der diagnostischen Güte zum Referenzkriterium Dysphagie auch die diagnostische Güte zu den Referenzkriterien Aspiration und Oralisierungseinschränkung dar. Bei diesen Ergebnissen ist zu beachten, dass sie nur eine Teilpopulation

der Patientinnen und Patienten mit Dysphagie darstellen, was als schwerwiegende Indirektheit beurteilt wurde. Zusätzlich wurden in 5 der 7 Indextests der Ablauf und / oder der Grenzwert retrospektiv festgelegt. Retrospektiv festgelegte Indextests wurden in diesem Evidenzbericht als sehr schwerwiegende Studienlimitationen bewertet.

Auffällig war, dass die diagnostische Güte zu keinem der dargestellten Indextests bisher repliziert wurde. Die Vergleichbarkeit der einzelnen Indextests wird zusätzlich durch die Unterschiede in der Durchführung und in den Kriterien, die zu einer Diagnose führten, erschwert (siehe Tabelle 5). Gleiches galt für die verwendeten Referenztests, sodass die Ergebnisse zur diagnostischen Güte kaum untereinander vergleichbar sind. In Anbetracht der Studienlage wird deutlich, dass lediglich die Nutzung von bildgebenden Instrumenten (FEES und VFSS) als Standard für die Diagnose der Dysphagie aufgefasst wird, nicht aber die zusätzlich benötigte einheitliche visuell perzeptive Bewertungsskala [43,44].

Ein weiterer Aspekt, der die Vergleichbarkeit der Ergebnisse der Studien und die Übertragbarkeit auf die deutsche Versorgungssituation erschwert ist, dass die Schweregrade der Kopf-Hals-Tumore uneinheitlich berichtet wurden. Zusätzlich bleibt für die Studien Horii 2021 und Carnaby 2014 unklar, ob alle Patientinnen und Patienten mit Anamnese einer Dysphagie ausgeschlossen wurden.

Ein Bedenken bei Schluckuntersuchungen ohne Bildgebung ist, dass Patientinnen und Patienten mit stillen Aspirationen unentdeckt bleiben könnten [45,46]. Aspirationen und somit auch stille Aspirationen gelten als Risikofaktor für die Entwicklung einer Aspirationspneumonie [5]. In keiner der vorliegenden Studien wurde berichtet, wie hoch der Anteil der Patientinnen und Patienten mit stiller Aspiration an den Patientinnen und Patienten mit falsch-negativem Ergebnis war. Somit ist anhand der Ergebnisse dieses Evidenzberichts zu diesen Bedenken keine Aussage möglich.

## 7 Literatur

1. Bundestag. Gesetz für eine bessere Versorgung durch Digitalisierung und Innovation (Digitale-Versorgung-Gesetz – DVG). Bundesgesetzblatt Teil 1 2019; (49): 2562-2584.
2. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften. S3-Leitlinie Oropharyngeale Dysphagie infolge einer Kopf-Hals-Tumor-Erkrankung - Diagnostik und Therapie [online]. 2018 [Zugriff: 07.02.2023]. URL: <https://register.awmf.org/de/leitlinien/detail/049-016>.
3. Schönemann H, Brożek J, Guyatt G et al. GRADE Handbook [online]. 2013 [Zugriff: 19.02.2021]. URL: <https://gdt.grade.pro.org/app/handbook/handbook.html>.
4. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden; Version 6.1 [online]. 2022 [Zugriff: 27.01.2022]. URL: <https://www.iqwig.de/methoden/allgemeine-methoden-v6-1.pdf>.
5. Baijens LWJ, Walshe M, Aaltonen LM et al. European white paper: oropharyngeal dysphagia in head and neck cancer. Eur Arch Otorhinolaryngol 2021; 278(2): 577-616. <https://dx.doi.org/10.1007/s00405-020-06507-5>.
6. ICH Expert Working Group. ICH harmonised tripartite guideline: structure and content of clinical study reports; E3 [online]. 1995 [Zugriff: 19.02.2021]. URL: [https://database.ich.org/sites/default/files/E3\\_Guideline.pdf](https://database.ich.org/sites/default/files/E3_Guideline.pdf).
7. Moher D, Hopewell S, Schulz KF et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010; 340: c869. <https://dx.doi.org/10.1136/bmj.c869>.
8. Des Jarlais DC, Lyles C, Crepaz N et al. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health 2004; 94(3): 361-366. <https://dx.doi.org/10.2105/ajph.94.3.361>.
9. Von Elm E, Altman DG, Egger M et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med 2007; 147(8): 573-577. <https://dx.doi.org/10.7326/0003-4819-147-8-200710160-00010>.
10. Bossuyt PM, Reitsma JB, Bruns DE et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Ann Intern Med 2003; 138(1): W1-12. <https://dx.doi.org/10.7326/0003-4819-138-1-200301070-00012-w1>.
11. Waffenschmidt S, Navarro-Ruan T, Hobson N et al. Development and validation of study filters for identifying controlled non-randomized studies in PubMed and Ovid MEDLINE. Res Synth Methods 2020; 11(5): 617-626. <https://dx.doi.org/10.1002/jrsm.1425>.

12. Guyatt GH, Oxman AD, Santesso N et al. GRADE guidelines: 12. Preparing summary of findings tables—binary outcomes. *J Clin Epidemiol* 2013; 66(2): 158-172.  
<https://dx.doi.org/10.1016/j.jclinepi.2012.01.012>.
13. Newcombe RG. Interval estimation for the difference between independent proportions: comparison of eleven methods. *Stat Med* 1998; 17(8): 873-890.  
[https://dx.doi.org/10.1002/\(sici\)1097-0258\(19980430\)17:8<873::aid-sim779>3.0.co;2-i](https://dx.doi.org/10.1002/(sici)1097-0258(19980430)17:8<873::aid-sim779>3.0.co;2-i).
14. Martín Andrés A, Silva Mato A. Choosing the optimal unconditioned test for comparing two independent proportions. *Comput Stat Data Anal* 1994; 17(5): 555-574.  
[https://dx.doi.org/10.1016/0167-9473\(94\)90148-1](https://dx.doi.org/10.1016/0167-9473(94)90148-1).
15. Sutton AJ, Abrams KR, Jones DR et al. *Methods for meta-analysis in medical research*. Chichester: Wiley; 2000.
16. Veroniki AA, Jackson D, Viechtbauer W et al. Recommendations for quantifying the uncertainty in the summary intervention effect and estimating the between-study heterogeneity variance in random-effects meta-analysis. *Cochrane Database Syst Rev* 2015; (10 Suppl 1): 25-27.
17. Guyatt GH, Oxman AD, Kunz R et al. GRADE guidelines: 7. Rating the quality of evidence—inconsistency. *J Clin Epidemiol* 2011; 64(12): 1294-1302.  
<https://dx.doi.org/10.1016/j.jclinepi.2011.03.017>.
18. Balshem H, Helfand M, Schunemann HJ et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol* 2011; 64(4): 401-406.  
<https://dx.doi.org/10.1016/j.jclinepi.2010.07.015>.
19. Meerpohl JJ, Langer G, Perleth M et al. GRADE-Leitlinien: 3. Bewertung der Qualität der Evidenz (Vertrauen in die Effektschätzer). *Z Evid Fortbild Qual Gesundheitswes* 2012; 106(6): 449-456. <https://dx.doi.org/10.1016/j.zefq.2012.06.013>.
20. Guyatt GH, Oxman AD, Vist G et al. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). *J Clin Epidemiol* 2011; 64(4): 407-415.  
<https://dx.doi.org/10.1016/j.jclinepi.2010.07.017>.
21. Guyatt GH, Oxman AD, Kunz R et al. GRADE guidelines: 8. Rating the quality of evidence—indirectness. *J Clin Epidemiol* 2011; 64(12): 1303-1310.  
<https://dx.doi.org/10.1016/j.jclinepi.2011.04.014>.
22. Guyatt GH, Oxman AD, Montori V et al. GRADE guidelines: 5. Rating the quality of evidence—publication bias. *J Clin Epidemiol* 2011; 64(12): 1277-1282.  
<https://dx.doi.org/10.1016/j.jclinepi.2011.01.011>.
23. Guyatt GH, Oxman AD, Kunz R et al. GRADE guidelines 6. Rating the quality of evidence—imprecision. *J Clin Epidemiol* 2011; 64(12): 1283-1293.  
<https://dx.doi.org/10.1016/j.jclinepi.2011.01.012>.

24. Guyatt GH, Oxman AD, Sultan S et al. GRADE guidelines: 9. Rating up the quality of evidence. *J Clin Epidemiol* 2011; 64(12): 1311-1316.  
<https://dx.doi.org/10.1016/j.jclinepi.2011.06.004>.
25. Whiting PF, Rutjes AW, Westwood ME et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; 155(8): 529-536.  
<https://dx.doi.org/10.7326/0003-4819-155-8-201110180-00009>.
26. Whiting PF, Rutjes AW, Westwood ME et al. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. *J Clin Epidemiol* 2013; 66(10): 1093-1104.  
<https://dx.doi.org/10.1016/j.jclinepi.2013.05.014>.
27. Leemis LM, Trivedi KS. A Comparison of Approximate Interval Estimators for the Bernoulli Parameter. *Am Statistn* 1996; 50(1): 63-68. <https://dx.doi.org/10.2307/2685046>.
28. Reitsma JB, Glas AS, Rutjes AW et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005; 58(10): 982-990. <https://dx.doi.org/10.1016/j.jclinepi.2005.02.022>.
29. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *J Clin Epidemiol* 2006; 59(12): 1331-1332; author reply 1332-1333. <https://dx.doi.org/10.1016/j.jclinepi.2006.06.011>.
30. Menke J. Bivariate random-effects meta-analysis of sensitivity and specificity with SAS PROC GLIMMIX. *Methods Inf Med* 2010; 49(1): 54-62, 62-54.  
<https://dx.doi.org/10.3414/ME09-01-0001>.
31. Hotelling H. The generalization of student's ratio. *Ann Math Stat* 1931; 2(3): 360-378.  
<https://dx.doi.org/10.1214/aoms/1177732979>.
32. Schünemann HJ, Mustafa RA, Brozek J et al. GRADE guidelines: 21 part 1. Study design, risk of bias, and indirectness in rating the certainty across a body of evidence for test accuracy. *J Clin Epidemiol* 2020; 122: 129-141.  
<https://dx.doi.org/10.1016/j.jclinepi.2019.12.020>.
33. Schünemann HJ, Mustafa RA, Brozek J et al. GRADE guidelines: 21 part 2. Test accuracy: inconsistency, imprecision, publication bias, and other domains for rating the certainty of evidence and presenting it in evidence profiles and summary of findings tables. *J Clin Epidemiol* 2020; 122: 142-152. <https://dx.doi.org/10.1016/j.jclinepi.2019.12.021>.
34. Hey C, Goeze A, Sader R et al. FraMaDySc: dysphagia screening for patients after surgery for head and neck cancer. *Eur Arch Otorhinolaryngol* 2023; 280(5): 2585-2592.  
<https://dx.doi.org/10.1007/s00405-023-07865-6>.
35. Hey C, Lange BP, Eberle S et al. Water swallow screening test for patients after surgery for head and neck cancer: early identification of dysphagia, aspiration and limitations of oral intake. *Anticancer Res* 2013; 33(9): 4017-4021.

36. Horii N, Hasegawa Y, Sakuramoto-Sadakane A et al. Validity of a dysphagia screening test following resection for head and neck cancer. *Ir J Med Sci* 2021; 190(1): 67-77.  
<https://dx.doi.org/10.1007/s11845-020-02286-4>.
37. Horii N, Hasegawa Y, Sakuramoto-Sadakane A et al. Correction to: Validity of a dysphagia screening test following resection for head and neck cancer. *Ir J Med Sci* 2021; 190(2): 861.  
<https://dx.doi.org/10.1007/s11845-020-02373-6>.
38. Patterson JM, Hildreth A, McColl E et al. The clinical application of the 100mL water swallow test in head and neck cancer. *Oral Oncol* 2011; 47(3): 180-184.  
<https://dx.doi.org/10.1016/j.oraloncology.2010.11.020>.
39. Carnaby GD, Crary MA. Development and validation of a cancer-specific swallowing assessment tool: MASA-C. *Support Care Cancer* 2014; 22(3): 595-602.  
<https://dx.doi.org/10.1007/s00520-013-2011-4>.
40. Hey C, Lange BP, Aere C et al. Predictability of oral and laryngopharyngeal function for aspiration and limitation of oral intake in patients after surgery for head and neck cancer. *Anticancer Res* 2013; 33(8): 3347-3353.
41. Omori F, Fujiu-Kurachi M, Wada K et al. Development of a Remote Examination of Deglutition Based on Consensus Surveys of Clinicians (Part II): Reliability and Validity in Healthy Elderly Individuals and Oral Cancer Patients. *Dysphagia* 2023; 38(3): 896-911.  
<https://dx.doi.org/10.1007/s00455-022-10514-5>.
42. Mann G, Hankey GJ, Cameron D. Swallowing disorders following acute stroke: prevalence and diagnostic accuracy. *Cerebrovasc Dis* 2000; 10(5): 380-386.  
<https://dx.doi.org/10.1159/000016094>.
43. Swan K, Cordier R, Brown T et al. Psychometric Properties of Visuoperceptual Measures of Videofluoroscopic and Fibre-Endoscopic Evaluations of Swallowing: A Systematic Review. *Dysphagia* 2019; 34(1): 2-33. <https://dx.doi.org/10.1007/s00455-018-9918-3>.
44. Speyer R, Cordier R, Farneti D et al. White Paper by the European Society for Swallowing Disorders: Screening and Non-instrumental Assessment for Dysphagia in Adults. *Dysphagia* 2022; 37(2): 333-349. <https://dx.doi.org/10.1007/s00455-021-10283-7>.
45. Boaden E, Burnell J, Hives L et al. Screening for aspiration risk associated with dysphagia in acute stroke. *Cochrane Database Syst Rev* 2021; 10: CD012679.  
<https://dx.doi.org/10.1002/14651858.CD012679.pub2>.
46. Brodsky MB, Suiter DM, Gonzalez-Fernandez M et al. Screening Accuracy for Aspiration Using Bedside Water Swallow Tests: A Systematic Review and Meta-Analysis. *Chest* 2016; 150(1): 148-163. <https://dx.doi.org/10.1016/j.chest.2016.03.059>.

47. Wong SS, Wilczynski NL, Haynes RB. Comparison of top-performing search strategies for detecting clinically sound treatment studies and systematic reviews in MEDLINE and EMBASE. *J Med Libr Assoc* 2006; 94(4): 451-455.

## 8 Studienlisten

### 8.1 Liste der identifizierten systematischen Übersichten

1. Dos Santos KW, da Cunha Rodrigues E, Rech RS et al. Using Voice Change as an Indicator of Dysphagia: A Systematic Review. *Dysphagia* 2022; 37(4): 736-748.  
<https://dx.doi.org/10.1007/s00455-021-10319-y>.

### 8.2 Liste der ausgeschlossenen Publikationen mit Ausschlussgründen

#### Nicht DN1

1. Bartlett RS, Kenz MK, Wayment HA et al. Correlation Between EAT-10 and Aspiration Risk Differs by Dysphagia Etiology. *Dysphagia* 2022; 37(1): 11-20.  
<https://dx.doi.org/10.1007/s00455-021-10244-0>.

2. Florie M, Baijens L, Kremer B et al. Relationship between swallow-specific quality of life and fiber-optic endoscopic evaluation of swallowing findings in patients with head and neck cancer. *Head Neck* 2016; 38 Suppl 1: E1848-1856. <https://dx.doi.org/10.1002/hed.24333>.

3. Florie M, Pilz W, Kremer B et al. EAT-10 Scores and Fiberoptic Endoscopic Evaluation of Swallowing in Head and Neck Cancer Patients. *Laryngoscope* 2021; 131(1): E45-E51.  
<https://dx.doi.org/10.1002/lary.28626>.

4. Lechien JR, Cavelier G, Thill MP et al. Validity and reliability of a French version of M.D. Anderson Dysphagia Inventory. *Eur Arch Otorhinolaryngol* 2020; 277(11): 3111-3119.  
<https://dx.doi.org/10.1007/s00405-020-06100-w>.

5. Lehner U, Zaretsky E, Goeze A et al. [Pre-treatment dysphagia in head-and-neck cancer patients]. *HNO* 2022; 70(7): 533-539. <https://dx.doi.org/10.1007/s00106-021-01128-8>.

6. Mandysova P, Ehler E, Skvrnakova J et al. Development of the Brief Bedside Dysphagia Screening Test - Revised: a Cross-Sectional Czech Study. *Acta Medica (Hradec Kralove)* 2015; 58(2): 49-55. <https://dx.doi.org/10.14712/18059694.2015.93>.

7. Printza A, Kyrgidis A, Pavlidou E et al. Reliability and validity of the Eating Assessment Tool-10 (Greek adaptation) in neurogenic and head and neck cancer-related oropharyngeal dysphagia. *Eur Arch Otorhinolaryngol* 2018; 275(7): 1861-1868.  
<https://dx.doi.org/10.1007/s00405-018-5001-9>.

#### Nicht DN2

1. Rahayu Rarasati R, Herawati Juniati S, Yusuf M. Conformity between modified Swallowing Outcome After Laryngectomy questionnaire and Flexible Endoscopic Evaluation of Swallowing in laryngeal carcinoma patients after total laryngectomy. *Journal of Public Health Research* 2022; 11(4): 22799036221127624.  
<https://dx.doi.org/10.1177/22799036221127624>.

### **Nicht DN3**

1. Carnaby GD, Madhavan A, Barikroo A et al. Change in Spontaneous Swallowing Frequency in HNC Patients Undergoing C/RT. *Otolaryngol Head Neck Surg* 2022; 166(4): 727-733. <https://dx.doi.org/10.1177/01945998211020744>.
2. Kendall KA, Kosek SR, Tanner K. Quality-of-life scores compared to objective measures of swallowing after oropharyngeal chemoradiation. *Laryngoscope* 2014; 124(3): 682-687. <https://dx.doi.org/10.1002/lary.24344>.

### **Nicht DN4**

1. Alvarez-Marcos C, Benito AV, Fernandez AG et al. Asymptomatic swallowing disorders may be present in individuals with laryngeal and hypopharyngeal cancer treated with chemo-radiotherapy. *Eur Arch Otorhinolaryngol* 2022; 279(2): 995-1001. <https://dx.doi.org/10.1007/s00405-021-06861-y>.
2. Andrade MS, Goncalves AN, Guedes RL et al. Correlation between swallowing-related quality of life and videofluoroscopy after head and neck cancer treatment. *Codas* 2017; 29(1): e20150175. <https://dx.doi.org/10.1590/2317-1782/20172015175>.
3. Arrese LC, Carrau R, Plowman EK. Relationship Between the Eating Assessment Tool-10 and Objective Clinical Ratings of Swallowing Function in Individuals with Head and Neck Cancer. *Dysphagia* 2017; 32(1): 83-89. <https://dx.doi.org/10.1007/s00455-016-9741-7>.
4. Arrese LC, Schieve HJ, Graham JM et al. Relationship between oral intake, patient perceived swallowing impairment, and objective videofluoroscopic measures of swallowing in patients with head and neck cancer. *Head Neck* 2019; 41(4): 1016-1023. <https://dx.doi.org/10.1002/hed.25542>.
5. Bofill-Soler N, Guillen-Sola A, Marco E et al. Is EAT-10 Useful to Assess Swallowing during the Chemo-Radiotherapy Phase in Patients with Head and Neck Cancer? A Pilot Study. *Ann Otol Rhinol Laryngol* 2021; 130(7): 689-698. <https://dx.doi.org/10.1177/0003489420966625>.
6. da Silva GM, Portas J, Lopez RVM et al. Study of Dysphagia in Patients with Advanced Oropharyngeal Cancer Subjected to an Organ Preservation Protocol Based on Concomitant Radiotherapy and Chemotherapy. *Asian Pac J Cancer Prev* 2019; 20(3): 977-982. <https://dx.doi.org/10.31557/apjcp.2019.20.3.977>.
7. Gupta DK, Singh SK, Sabarigirish K et al. Evaluation of swallowing in patients with T3/T4 oral squamous cell carcinoma. *Medical Journal Armed Forces India* 2023; 79(2): 181-188. <https://dx.doi.org/10.1016/j.mjafi.2021.10.002>.
8. Hedstrom J, Tuomi L, Finizia C et al. Correlations Between Patient-Reported Dysphagia Screening and Penetration-Aspiration Scores in Head and Neck Cancer Patients Post-oncological Treatment. *Dysphagia* 2018; 33(2): 206-215. <https://dx.doi.org/10.1007/s00455-017-9847-6>.

9. Ishii R, Kato K, Ohkoshi A et al. Simultaneous evaluation of symptoms, swallowing functions, and patient-reported swallowing difficulties and their correlations with ingestion status during definitive chemoradiotherapy for oropharyngeal and hypopharyngeal cancer. *Support Care Cancer* 2021; 29(2): 955-964. <https://dx.doi.org/10.1007/s00520-020-05570-2>.
10. Nevens D, Goeleven A, Duprez F et al. Does the total dysphagia risk score correlate with swallowing function examined by videofluoroscopy? *Br J Radiol* 2018; 91(1083): 20170714. <https://dx.doi.org/10.1259/bjr.20170714>.
11. Nevens D, Goeleven A, Duprez F et al. Correlation of Patient- and Physician-Scored Dysphagia with Videofluoroscopies in Patients Treated with Radiotherapy for Head and Neck Cancer. *Dysphagia* 2018; 33(5): 684-690. <https://dx.doi.org/10.1007/s00455-018-9888-5>.
12. Pedersen A, Wilson J, McColl E et al. Swallowing outcome measures in head and neck cancer--How do they compare? *Oral Oncol* 2016; 52: 104-108. <https://dx.doi.org/10.1016/j.oraloncology.2015.10.015>.
13. Pizzorni N, Schindler A, Fantini M et al. Relationship between swallowing-related quality of life and fiberoptic endoscopic evaluation of swallowing in patients who underwent open partial horizontal laryngectomy. *Eur Arch Otorhinolaryngol* 2018; 275(4): 973-985. <https://dx.doi.org/10.1007/s00405-018-4888-5>.
14. Wishart LR, Harris GB, Cassim N et al. Association Between Objective Ratings of Swallowing and Dysphagia-Specific Quality of Life in Patients Receiving (Chemo)radiotherapy for Oropharyngeal Cancer. *Dysphagia* 2022; 37(4): 1014-1021. <https://dx.doi.org/10.1007/s00455-021-10364-7>.
15. Yang CJ, Roh JL, Choi KH et al. Pretreatment Dysphagia Inventory and videofluorographic swallowing study as prognostic indicators of early survival outcomes in head and neck cancer. *Cancer* 2015; 121(10): 1588-1598. <https://dx.doi.org/10.1002/cncr.29245>.
16. Zaretsky E, Steinbach-Hundt S, Pluschinski P et al. [Validation of the German version of Eating Assessment Tool for head and neck cancer patients]. *Laryngorhinootologie* 2018; 97(7): 480-486. <https://dx.doi.org/10.1055/a-0596-7780>.

## Anhang A Kriterien des Verzerrungspotenzials

Tabelle 17: Kriterien des Verzerrungspotenzials nach QUADAS-2 zum Referenzkriterium Dysphagie

Studie	Personenselektion	Indextest	Referenzstandard	Personenfluss und zeitlicher Ablauf
Carnaby 2014	unklar	hoch	unklar	hoch
Hey 2023	unklar	hoch <sup>a</sup> / niedrig <sup>b</sup>	unklar	niedrig
Hey 2013	unklar	hoch	unklar	niedrig
a. Bewertung für den Indextest FraMaDySc b. Bewertung für den Indextest 51 ml WST FraMaDySc: Frankfurt/Marburg Dysphagia Screening; QUADAS: Quality Assessment of Diagnostic Accuracy Studies; WST: Wasserschlucktest				

Tabelle 18: Kriterien des Verzerrungspotenzials nach QUADAS-2 zum Referenzkriterium Aspiration

Studie	Personenselektion	Indextest	Referenzstandard	Personenfluss und zeitlicher Ablauf
Hey 2023	unklar	hoch <sup>a</sup> / niedrig <sup>b</sup>	unklar	niedrig
Hey 2013	unklar	hoch	unklar	niedrig
Horii 2021	hoch	hoch	unklar	hoch
Patterson 2011	niedrig	unklar	hoch	hoch
a. Bewertung für den Indextest FraMaDySc b. Bewertung für den Indextest 51 ml WST FraMaDySc: Frankfurt/Marburg Dysphagia Screening; QUADAS: Quality Assessment of Diagnostic Accuracy Studies; WST: Wasserschlucktest				

Tabelle 19: Kriterien des Verzerrungspotenzials nach QUADAS-2 zum Referenzkriterium Oralisierungseinschränkung

<b>Studie</b>	<b>Personenselektion</b>	<b>Indextest</b>	<b>Referenzstandard</b>	<b>Personenfluss und zeitlicher Ablauf</b>
Hey 2023	unklar	hoch <sup>a</sup> / niedrig <sup>b</sup>	unklar	niedrig
Hey 2013	unklar	hoch	unklar	niedrig
a. Bewertung für den Indextest FraMaDySc b. Bewertung für den Indextest 51 ml WST FraMaDySc: Frankfurt/Marburg Dysphagia Screening; QUADAS: Quality Assessment of Diagnostic Accuracy Studies; WST: Wasserschlucktest				

**Anhang B Autorenanfragen**

Tabelle 20: Übersicht über Autorenanfragen

Studie	Inhalt der Anfrage	Antwort eingegangen ja / nein	Inhalt der Antwort
Carnaby 2014	<ul style="list-style-type: none"> <li>▪ Bitte um Übermittlung der Vierfeldertafel-Daten für den Indextest (MASA-C) für das Referenzkriterium der Dysphagie</li> <li>▪ Bitte um Erklären der Diskrepanz zwischen der Anzahl an Patientinnen und Patienten mit richtig-positiver erkannter Dysphagie unter dem Ergebnis der Zielgröße „diagnostic yield“ und dem Ergebnis der VFFS auf den Seiten 598-599</li> </ul>	nein	–
Hey 2013	<ul style="list-style-type: none"> <li>▪ Bitte um Übermittlung der Vierfeldertafel-Daten für die Indextests (Kombination 1 und 2) für das Referenzkriterium der Dysphagie</li> <li>▪ Bitte um Erklären der Diskrepanzen zwischen den Vierfeldertafel-Daten (Summe aus richtig-positiv und falsch-negativ) in Tabelle V zu den Kombinationen 1 und 2 im Vergleich zu der in Tabelle III berichteten Anzahl der Patientinnen und Patienten mit Aspiration bzw. mit Oralisierungseinschränkung</li> </ul>	ja	<ul style="list-style-type: none"> <li>▪ Vierfeldertafel-Daten für die Indextests (Kombination 1 und 2) für das Referenzkriterium Dysphagie</li> <li>▪ korrigierte Vierfeldertafel-Daten zu Tabelle V</li> </ul>
Hey 2023	<ul style="list-style-type: none"> <li>▪ Umfasst die Studienpopulation von Hey 2023, auch die Patientinnen und Patienten, welche in einer Publikation von Hey et al. aus dem Jahr 2013 [35] beschrieben werden?</li> <li>▪ Bitte um Erklären der Diskrepanzen zwischen den Vierfeldertafel-Daten (Summe aus richtig-positiv und falsch-negativ) in Tabelle 4 für den Indextest FraMaDySc im Vergleich zu der im Text berichteten Anzahl der Patientinnen und Patienten mit Aspiration bzw. mit Oralisierungseinschränkung</li> </ul>	ja	<ul style="list-style-type: none"> <li>▪ ja, die Patientinnen und Patienten aus der Publikation von Hey et al. aus dem Jahr 2013 [35] gehen in die Ergebnisse der Studie Hey 2023 ein</li> <li>▪ korrigierte Vierfeldertafel-Daten zu Tabelle 4</li> </ul>
Omori 2022	<ul style="list-style-type: none"> <li>▪ Bitte um Übermittlung der Vierfeldertafel-Daten für den Indextest (RED) für das Referenzkriterium der Dysphagie</li> <li>▪ Bitte um Übermittlung der Vierfeldertafel-Daten für den des Indextest (RED) für das Referenzkriterium der Aspiration</li> <li>▪ Bitte um Erklären der Diskrepanz zwischen der Anzahl an Patientinnen und Patienten ohne Dysphagie in Tabelle 5 und den Ergebnissen der instrumentellen Diagnostik (Referenztest) auf Seite 902.</li> <li>▪ Bitte um Erklären der Diskrepanz zwischen der Anzahl an Patientinnen und Patienten mit Aspiration in Tabelle 5 und den Ergebnissen der instrumentellen Diagnostik (Referenztest) auf Seite 902.</li> </ul>	nein	–
RED: remote examination of deglutition (klinische Schluckuntersuchung); VFSS: videofluoroskopische Schluckuntersuchung			

## Anhang C Suchstrategien

### C.1 Bibliografische Datenbanken

#### Suche nach systematischen Übersichten

##### 1. MEDLINE

*Suchoberfläche: Ovid*

- Ovid MEDLINE(R) ALL 1946 to March 31, 2023

Es wurde folgender Filter übernommen:

- Systematische Übersicht: Wong [47] – High specificity strategy

#	Searches
1	exp Deglutition Disorders/
2	(swallow* or deglutit* or dysphag* or xerostom* or saliva* or dryness* or aspiration* or ((pharyn* or oropharyn*) adj3 (disturbance* or disorder* or difficult* or dysfunction* or impair* or condition* or abnormal* or damag* or injur*))).ti,ab.
3	or/1-2
4	Fluoroscopy/
5	(videofluoroscop* or (fiberoptic adj1 endoscopic*) or FEES or VFSS).ti,ab.
6	(videofluorograph* or videoendoscop* or (barium adj1 swallow*)).ti,ab.
7	((eat* adj1 assessment adj1 tool*) or EAT?10).ti,ab.
8	((dysphagia adj1 handicap* adj1 index*) or DHI).ti,ab.
9	((sydney* adj1 swallow* adj1 questionnaire*) or SSQ).ti,ab.
10	((anderson adj1 dysphagia adj1 inventor*) or MDADI).ti,ab.
11	((european adj1 organisation* adj8 quality adj5 questionnaire*) or EORTC?QLQ).ti,ab.
12	(dysphagia* adj3 (questionnaire* or (self-report* adj3 assess*))).ti,ab.
13	or/4-12
14	and/3,13
15	Cochrane database of systematic reviews.jn.
16	(search or MEDLINE or systematic review).tw.
17	meta analysis.pt.
18	or/15-17
19	18 not (exp animals/ not humans.sh.)
20	and/14,19
21	20 and (english or german or multilingual or undetermined).lg.
22	..l/ 21 yr=2018-Current

## 2. International HTA Database

*Suchoberfläche: INAHTA*

#	Searches
1	Deglutition Disorders[mhe]
2	(swallow* or deglutit* or dysphag* or xerostom* or saliva* or dryness* or aspiration* or ((pharyn* or oropharyn*) AND (disturbance* or disorder* or difficult* or dysfunction* or impair* or condition* or abnormal* or damag* or injur*))) [title] OR (swallow* or deglutit* or dysphag* or xerostom* or saliva* or dryness* or aspiration* or ((pharyn* or oropharyn*) AND (disturbance* or disorder* or difficult* or dysfunction* or impair* or condition* or abnormal* or damag* or injur*))) [abs]
3	#2 OR #1
4	(*) FROM 2018 TO 2023
5	#4 AND #3

### Suche nach Primärstudien

#### 1. MEDLINE

*Suchoberfläche: Ovid*

- Ovid MEDLINE(R) ALL 1946 to April 05, 2023

#	Searches
1	exp "Head and Neck Neoplasms"/
2	((head or neck or upper aerodigestive tract* or facial or oral or mouth or lip* or tongue or throat or laryn* or pharyn* or hypopharyn* or oropharyn* or nasopharyn* or paranasal or nasal or nose or salivary or parotid or glotti* or subglotti* or supraglotti* or oesophag* or esophag*) adj3 (carcinom* or cancer* or tumo?r*)).ti,ab.
3	or/1-2
4	exp Deglutition Disorders/
5	(swallow* or deglutit* or dysphag* or xerostom* or saliva* or dryness* or aspiration* or ((pharyn* or oropharyn*) adj3 (disturbance* or disorder* or difficult* or dysfunction* or impair* or condition* or abnormal* or damag* or injur*))).ti,ab.
6	or/4-5
7	Fluoroscopy/
8	(videofluoroscop* or (fiberoptic adj1 endoscopic*) or FEES or VFSS).ti,ab.
9	(videofluorograph* or videoendoscop* or (barium adj1 swallow*)).ti,ab.
10	((swallow* or dysphagia*) adj3 (water* or assessment or test* or screen* or examination*)).ti,ab.
11	((eat* adj1 assessment adj1 tool*) or EAT?10).ti,ab.
12	((dysphagia adj1 handicap* adj1 index*) or DHI).ti,ab.
13	((sydney* adj1 swallow* adj1 questionnaire*) or SSQ).ti,ab.
14	((anderson adj1 dysphagia adj1 inventor*) or MDADI).ti,ab.
15	((european adj1 organisation* adj8 quality adj5 questionnaire*) or EORTC?QLQ).ti,ab.
16	(dysphagia* adj3 (questionnaire* or (self-report* adj3 assess*))).ti,ab.
17	or/7-16

#	Searches
18	and/3,6,17
19	(animals/ not humans/) or comment/ or editorial/ or exp review/ or meta analysis/ or consensus/ or exp guideline/
20	hi.fs. or case report.mp.
21	or/19-20
22	18 not 21
23	22 and (english or german or multilingual or undetermined).lg.
24	23 and 2011:3000.(dt).

## 2. The Cochrane Library

Suchoberfläche: Wiley

- Cochrane Central Register of Controlled Trials: Issue 4 of 12, April 2023

#	Searches
#1	[mh "Head and Neck Neoplasms"]
#2	((head or neck or upper aerodigestive tract* or facial or oral or mouth or lip* or tongue or throat or larynx* or pharynx* or hypopharynx* or oropharynx* or nasopharynx* or paranasal or nasal or nose or salivary or parotid or glotti* or subglotti* or supraglotti* or oesophag* or esophag*) NEAR/3 (carcinom* or cancer* or tumor*)):ti,ab
#3	#1 or #2
#4	[mh "Deglutition Disorders"]
#5	(swallow* or deglutit* or dysphag* or xerostom* or saliva* or dryness* or aspiration* or ((pharynx* or oropharynx*) NEAR/3 (disturbance* or disorder* or difficult* or dysfunction* or impair* or condition* or abnormal* or damag* or injur*)):ti,ab
#6	#4 or #5
#7	[mh ^"Fluoroscopy"]
#8	(videofluoroscop* or (fiberoptic NEAR/1 endoscopic*) or FEES or VFSS):ti,ab
#9	(videofluorograph* or videoendoscop* or (barium NEAR/1 swallow*)):ti,ab
#10	((swallow* or dysphagia*) NEAR/3 (water* or assessment or test* or screen* or examination*)):ti,ab
#11	((eat* NEAR/1 assessment NEAR/1 tool*) or EAT?10):ti,ab
#12	((dysphagia NEAR/1 handicap* NEAR/1 index*) or DHI):ti,ab
#13	((sydney* NEAR/1 swallow* NEAR/1 questionnaire*) or SSQ):ti,ab
#14	((anderson NEAR/1 dysphagia NEAR/1 inventor*) or MDADI):ti,ab
#15	((european NEAR/1 organisation* NEAR/8 quality NEAR/5 questionnaire*) or EORTC?QLQ):ti,ab
#16	(dysphagia* NEAR/3 (questionnaire* or (self-report* NEAR/3 assess*)):ti,ab
#17	#7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16
#18	#3 AND #6 AND #17
#19	#18 NOT (*clinicaltrial*gov* or *trialssearch*who* or *clinicaltrialsregister*eu* or *anzctr*org*au* or *trialregister*nl* or *irct*ir* or *isrctn* or *controlled*trials*com* or *drks*de*):so

#	Searches
#20	#19 not ((language next (afr or ara or aze or bos or bul or car or cat or chi or cze or dan or dut or es or est or fin or fre or gre or heb or hrv or hun or ice or ira or ita or jpn or ko or kor or lit or nor or peo or per or pol or por or pt or rom or rum or rus or slo or slv or spa or srp or swe or tha or tur or ukr or urd or uzb)) not (language near/2 (en or eng or english or ger or german or mul or unknown)))
#21	#20 with Publication Year from 2011 to 2023, in Trials

## C.2 Studienregister

### ClinicalTrials.gov

**Anbieter: U.S. National Institutes of Health**

- URL: <http://www.clinicaltrials.gov>
- Eingabeoberfläche: Expert Search

Suchstrategie
( dysphagia OR swallow OR aspiration OR deglutition ) AND AREA[ResultsFirstSubmitDate] NOT MISSING AND ( videofluoroscopy OR fiberoptic endoscopic OR FEES OR VFSS OR water swallow test OR clinical swallow examination )

## 2. International Clinical Trials Registry Platform Search Portal

**Anbieter: World Health Organization**

- URL: <https://trialssearch.who.int>
- Eingabeoberfläche: Standard Search

Suchstrategie
( dysphagia OR swallow* OR aspiration OR deglut* ) AND ( videofluoroscopy OR fiberoptic endoscopic OR FEES OR VFSS OR water swallow test OR clinical swallow examination ) with results only