

Aus der Abteilung für Zahnerhaltung und Präventivmedizin
der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

Sensitivität und Spezifität intraoraler Röntgenaufnahmen zur Detektion von
Karies – eine systematische Metaanalyse und Übersichtsarbeit

zur Erlangung des akademischen Grades
Doctor medicinae dentariae (Dr. med. dent.)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Markus Tzschope

aus Bad Muskau

Datum der Promotion: 30.11.2023

Meiner Familie

Vorwort

Hiermit weise ich, Markus Tzschoppe, darauf hin, dass Teilergebnisse der hier vorliegenden Arbeit in folgender Publikation veröffentlicht wurden:

Schwendicke F, Tzschoppe M, Paris S (2015): Radiographic caries detection: A Systematic Review and Meta-Analysis, *Journal of Dentistry*, 43;924-9

Inhaltsverzeichnis

| | |
|---|-----------|
| Inhaltsverzeichnis | 4 |
| Tabellenverzeichnis | 5 |
| Abbildungsverzeichnis | 6 |
| Abstrakt (deutsch) | 7 |
| Abstract (english) | 8 |
| 1 Einleitung | 9 |
| 2 Material und Methoden | 17 |
| 2.1 Einschlusskriterien | 17 |
| 2.1.1 Population | 17 |
| 2.1.2 Indextest | 18 |
| 2.1.3 Referenztest | 19 |
| 2.1.4 Testergebnisse | 19 |
| 2.2 Informationsquellen | 19 |
| 2.3 Suche und Auswahl der Studien | 20 |
| 2.4 Datenextraktion | 20 |
| 2.5 Verzerrungspotenzial | 21 |
| 2.6 Synthese | 22 |
| 3 Ergebnisse | 24 |
| 3.1 Suchergebnisse | 24 |
| 3.2 Eingeschlossene Studien und Verzerrungspotenzial | 30 |
| 3.3 Datenauswertung | 30 |
| 3.4 Publikationsbias | 35 |
| 4 Diskussion | 39 |
| Literaturverzeichnis | 47 |
| Eidesstattliche Versicherung | 52 |
| Anteilerklärung an etwaigen erfolgten Publikationen | 53 |
| Lebenslauf/Vita | 54 |
| Komplette Publikationsliste | 55 |
| Danksagung | 56 |

Tabellenverzeichnis

Tabelle 2-1: QUADAS-2-Richtlinien: Verzerrungspotenzial und Bedenken hinsichtlich der Anwendbarkeit der Ergebnisse, 1 = niedriges Risiko, 2 = hohes Risiko, 3 = unklares Risiko.

Tabelle 3-1: Eingeschlossene Studien nach QUADAS-2-Richtlinien: Verzerrungspotenzial und Bedenken hinsichtlich der Anwendbarkeit der Ergebnisse, 1 = niedriges Risiko, 2 = hohes Risiko, 3 = unklares Risiko.

Tabelle 3-2: Sensitivität, Spezifität und Genauigkeit röntgenologischer Kariesdetektion für die verschiedenen Schwellenwerte und Zahnflächen.

Whiting P, Rutjes A, Westwood M, Mallett S, Deeks J, Reitsma J, Leeflang MGM, Sterne JAC, Bossuyt PMM, QUADAS-2 Group (2011): QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies, *Annals of Internal Medicine*, 155:529–36.

In Anlehnung an Whiting P, Rutjes A, Westwood M, Mallett S, Deeks J, Reitsma J, Leeflang MGM, Sterne JAC, Bossuyt PMM, QUADAS-2 Group (2011): QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies, *Annals of Internal Medicine*, 155:529–36.

Schwendicke F, Tzschoppe M, Paris S (2015): Radiographic caries detection: A Systematic Review and Meta-Analysis, *Journal of Dentistry*, 43:924-933.

Abbildungsverzeichnis

- Abbildung 1-1:** Genauigkeit, Richtigkeit und Präzision grafisch dargestellt. In Anlehnung an Barwick VJ; Prichard E (2011): Terminologie bei Analytischen Messungen. Eine Einführung in den VIM 3. 1. Aufl. Online verfügbar unter www.eurachem.org.
- Abbildung 1-2:** Die 4-Felder-Tafel. Meyer-Lückel H, Paris S, Ekstrand K (2012): Karies – Wissenschaft und klinische Praxis.
- Abbildung 3-1:** Der Suchprozess. Schwendicke F, Tzschoppe M, Paris S (2015): Radiographic caries detection: A Systematic Review and Meta-Analysis, *Journal of Dentistry*, 43;924-933.
- Abbildung 3-2:** ROC-Kurven. Schwendicke F, Tzschoppe M, Paris S (2015): Radiographic caries detection: A Systematic Review and Meta-Analysis, *Journal of Dentistry*, 43;924-933.
- Abbildung 3-3:** Funnel-Plots. Schwendicke F, Tzschoppe M, Paris S (2015): Radiographic caries detection: A Systematic Review and Meta-Analysis, *Journal of Dentistry*, 43;924-933.
- Abbildung 3-4:** Metaregressionsanalyse. Schwendicke F, Tzschoppe M, Paris S (2015): Radiographic caries detection: A Systematic Review and Meta-Analysis, *Journal of Dentistry*, 43;924-933.

Abstrakt (deutsch)

Zielsetzungen: Diese systematische Übersichtsarbeit zielte darauf ab, die Genauigkeit von Bissflügelröntgen zur Detektion von Karies zu bewerten. Daten: Studien, die Angaben über die Genauigkeit (Sensitivität/Spezifität) des röntgenologischen Nachweises natürlicher kariöser Läsionen unter klinischen oder In-vitro-Bedingungen beinhalteten, wurden eingeschlossen. Das Verzerrungspotenzial wurde mithilfe der QUADAS-2-Richtlinien bewertet. Die gepoolten Sensitivitäts- und Spezifitätswerte sowie das Quotenverhältnis (DOR) wurden mithilfe einer Metaanalyse mit zufälligen Effekten berechnet. Die Analysen wurden getrennt für okklusale und proximale Zahnflächen durchgeführt, wobei zwischen allen jeglicher Art von Läsionen, Dentinläsionen und kavitierten Läsionen weiter unterschieden wurde.

Quellen: Elektronische Datenbanken (Medline, Embase, Cochrane Central) und graue Literatur wurden systematisch durchsucht und ergänzt durch Querverweise aus Bibliografien. Studienauswahl: Von 947 identifizierten Artikeln wurden 442 im Volltext analysiert. 117 Studien (13.375 Zähne, 19.108 Oberflächen) wurden eingeschlossen, wobei die meisten von ihnen über bleibende Zähne berichteten und ein hohes Verzerrungspotenzial aufwiesen. Der Nachweis jeglicher Art von Läsionen (d. h. auch initialer) wies geringe Sensitivitäten (gepooltes DOR [95 % CI]: 0,24 [0,21/0,26] bis 0,42 [0,31/0,34]), aber moderate bis hohe Spezifitäten auf (0,70 [0,76/0,84] bis 0,97 [0,95/0,98]). Für Dentinläsionen waren die Sensitivitäten höher (von 0,36 [0,24/0,49] für proximale bis 0,56 [0,53/0,59] für okklusale Läsionen) und die Spezifitäten lagen zwischen 0,87 [0,85/0,89] und 0,95 [0,94/0,96]. Keine Studie berichtete über kavitierte okklusale Läsionen, während bei kavitierten proximalen Läsionen die Sensitivitäten über 0,60 anstiegen und die Spezifitäten hoch blieben (über 0,90).

Schlussfolgerungen: Die röntgenologische Kariesdetektion ist bei kavitierten proximalen Läsionen sehr genau und scheint auch für den Nachweis von Dentinläsionen geeignet zu sein. Zur Erkennung initialer Läsionen könnten sensitivere Detektionsmethoden in Populationen mit hohem Kariesrisiko und hoher Prävalenz in Betracht gezogen werden. Klinische Bedeutung: Die röntgenologische Kariesdetektion eignet sich besonders zur Erkennung von fortgeschritteneren Kariesläsionen und hat ein begrenztes Risiko für falsch positive Diagnosen. Für Gruppen mit hohem Kariesrisiko und hoher Prävalenz könnten alternative Nachweismethoden mit höherer Sensitivität zur Detektion initialer Läsionen in Betracht gezogen werden.

Schlüsselwörter: Bissflügel, Karies, Dental, Röntgen

Abstract (english)

Objectives: This systematic review aimed at evaluating the accuracy of radiographic caries detection for different lesions at different locations. Data: Studies reporting on the accuracy (sensitivity/specificity) of radiographic detection of natural primary caries lesions under clinical or in vitro conditions were included. Risk of bias was assessed using QUADAS-2. Pooled sensitivity, specificity and diagnostic odds ratios (DORs) were calculated using random-effects meta-analysis. Analyses were performed separately for occlusal and proximal lesions, with further discrimination between any kind of lesions, dentine lesions, and cavitated lesions.

Sources: Electronic databases (Medline, Embase, Cochrane Central) and grey literature were systematically searched, complemented by cross-referencing from bibliographies. Study selection: From 947 identified articles, 442 were analyzed full-text. 117 studies (13,375 teeth, 19,108 surfaces) were included, the majority of them reporting on permanent teeth and having high risk of bias. The detection of any kind (i. e. also initial) lesions had low sensitivities (pooled DOR [95 % CI]: 0.24 [0.21/0.26] to 0.42 [0.31/0.34]), but moderate to high specificities (0.70 [0.76/0.84] to 0.97 [0.95/0.98]). For dentine lesions, sensitivities were higher (from 0.36 [0.24/0.49] for proximal to 0.56 [0.53/0.59] for occlusal lesions), and specificities ranged between 0.87 [0.85/0.89] and 0.95 [0.94/0.96]. No studies reported on cavitated occlusal lesions, whilst for cavitated proximal lesions, sensitivities increased above 0.60, whilst specificities remained high (above 0.90).

Conclusions: Radiographic caries detection is highly accurate for cavitated proximal lesions, and seems also suitable to detect dentine caries lesions. For detecting initial lesions, more sensitive methods could be considered in population with high caries risk and prevalence. Clinical significance: Radiographic caries detection is especially suitable for detecting more advanced caries lesions, and has limited risks for false positive diagnoses. For groups with high caries risk and prevalence, alternative detection methods with higher sensitivity for initial lesions might be considered.

Keywords: Bitewing, Caries, Dental, Radiography

1 Einleitung

Das Screening nach kariösen Läsionen stellt einen zentralen Bestandteil der täglichen zahnärztlichen Praxis dar. Das Ziel ist dabei die Detektion in einem möglichst frühen Stadium, um schließlich non- oder mikro-invasive Behandlungsmaßnahmen einleiten zu können und somit wesentlich invasivere und letztlich auch teurere restaurative Therapieformen vermeiden zu können (Bader et al., 2001; Gordan et al., 2009; Rindal et al., 2010). Durch visuell-taktile Diagnostik ist es möglich, kariöse Läsionen an okklusalen, bukkalen und lingualen Flächen zu diagnostizieren. Jedoch stellt es sich schon deutlich schwieriger dar, insbesondere nicht kavitierte proximale kariöse Läsionen an jenen Prädilektionsstellen zu diagnostizieren, welche weder dem Auge des Zahnarztes zugänglich sind, noch durch dessen Sonde im Rahmen der visuell-taktilen Diagnostik erreicht werden können (Bader et al., 2001). Diese Läsionen lassen sich durch verschiedene zusätzliche Methoden diagnostizieren, wobei das (Bissflügel-)Röntgenverfahren hierbei am häufigsten zur Anwendung kommt (Rindal et al., 2010); pro Jahr und Praxis werden in Deutschland im Durchschnitt etwa 135 Panoramaschichtaufnahmen und 550 Zahnfilme angefertigt (Statistisches Jahrbuch der Bundeszahnärztekammer, 2013/14). Dabei werden kariöse Läsionen durch die infolge des Mineralverlusts zunehmende Radiotransluzenz der Zahnhartsubstanz sichtbar und somit schließlich detektierbar (also erkennbar) sowie diagnostizierbar (in Bezug auf Progressionsgrad einschätzbar) gemacht.

Das konventionelle Röntgenverfahren basiert auf der im Jahre 1895 von Wilhelm Conrad Röntgen entwickelten Technik. Auch wenn sich am Grundprinzip des Verfahrens bis heute nichts verändert hat, wurde es im Laufe der Zeit immer weiterentwickelt. So bieten die in den letzten Jahren entwickelten digitalen Röntgenverfahren, wie zum Beispiel jene beruhend auf Halbleitersensoren oder Lumineszenzfolien, gegenüber konventionellen Röntgenverfahren diverse Vorteile. Zu diesen Vorteilen zählen unter anderem das verringerte Auftreten von Darstellungsfehlern, eine erhöhte Belichtungstoleranz und eine bessere Umweltverträglichkeit durch den Wegfall der Nasschemie. Vor allem ist aber die Einsparung von Zeit und Dosis hervorzuheben. Somit werden in praxi zum einen effizientere Arbeitsabläufe ermöglicht, zum anderen wird dem Patienten ein gewisser Teil der Strahlendosis erspart (Kidd et al., 2003).

Der Vollständigkeit halber sei erwähnt, dass nunmehr auch alternative Diagnostikverfahren existieren, welche ohne ionisierende Strahlung auskommen. Hierzu zählen zum Beispiel auf Fluoreszenz basierende Verfahren. Diese wurden im Rahmen verschiedener Studien für höchst

sensitiv, jedoch weniger spezifisch befunden. Der Vorteil jener Verfahren im Vergleich zur röntgenologischen Kariesdiagnostik könnte letztlich zum einen darin bestehen, dass hierbei auf potenziell schädliche ionisierende Strahlung verzichtet wird. Zum anderen könnten die hier zur Anwendung kommenden Hilfsmittel auch leichter und effizienter in der täglichen zahnärztlichen Praxis anwendbar sein (Gimenez et al., 2013).

Im Hinblick auf die Auswahl unterschiedlicher zur Verfügung stehender Detektionsmethoden erscheint es sinnvoll, diese zu vergleichen beziehungsweise in ihrer Güte einschätzen zu können. Im Folgenden werden zunächst die drei klassischen Gütekriterien Objektivität, Reliabilität und Validität vorgestellt. Bei diesen drei Hauptkriterien geht es im weitesten Sinne immer um verschiedene Arten von Genauigkeit beziehungsweise die Einordnung der Genauigkeit in einen weiter gefassten Kontext. Testergebnisse sind dann genau, wenn sie sowohl richtig als auch präzise sind. Dabei gibt die Richtigkeit einer Messung Auskunft über die Annäherung eines Erwartungswertes an einen wahren Wert. Bei der Präzision handelt es sich um einen Aspekt zur Beurteilung der Qualität einer Messung. Dieser Gesichtspunkt beschreibt die Streubreite der Ergebnisse der zugrundeliegenden Messung (Abb. 1-1).

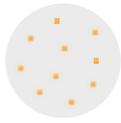
| | | | |
|---|---|--|---|
|  |  |  |  |
| Richtigkeit gut | Richtigkeit schlecht | Richtigkeit gut | Richtigkeit schlecht |
| Präzision gut | Präzision gut | Präzision schlecht | Präzision schlecht |
| Genauigkeit gut | Genauigkeit schlecht | Genauigkeit schlecht | Genauigkeit schlecht |

Abbildung 1-1: Genauigkeit, Richtigkeit und Präzision grafisch dargestellt. Es werden jeweils mehrere Schüsse mit einem fest eingespannten Gewehr auf eine Zielscheibe abgegeben (in Anlehnung an Barwick VJ et al., 2011).

Die Objektivität eines Tests ist die Unabhängigkeit der Testergebnisse von den Personen, die bei der Ergebnisfindung beteiligt sind. Die Reliabilität, beziehungsweise Zuverlässigkeit, stellt per definitionem denjenigen Anteil an Variabilität dar, welcher durch tatsächliche Unterschiede im zu messenden Merkmal besteht und somit nicht durch Messfehler verursacht werden kann. Ein hochreliaables Testverfahren ist demnach frei von Zufallsfehlern, sodass die Wiederholung einer Messung das gleiche Messergebnis erbringen kann. Bei der Reliabilität handelt es sich also um eine Art von Genauigkeit, die allerdings unabhängig davon ist, ob gemessen wird, was gemessen werden soll. Die Validität ist die Gültigkeit einer Messung. Sie beschreibt, ob gemessen wird, was

gemessen werden soll. Es werden dabei verschiedene Arten der Validität unterschieden. Hier wären als Wichtigste zu nennen:

- Die Konstruktvalidität. Sie liegt vor, wenn die Messung eines Konstrukts weder durch systematische Fehler noch durch andere Konstrukte verfälscht wird.
- Die Inhaltsvalidität. Sie bezeichnet in der multivariaten Statistik einen Teilaspekt der Konstruktvalidität und liegt vor, wenn die Messungen eines Konstrukts dessen Inhalt in all seinen Aspekten vollständig erfassen.
- Die Kriteriumsvalidität. Sie gibt an, inwieweit die zu validierende Messung eines Konstrukts mit der Messung eines externen Kriteriums, dem sogenannten Außenkriterium, von dem bekannt ist, dass es das Konstrukt valide erfasst bzw. dass es in einem validen kausalen Zusammenhang mit dem Konstrukt steht, übereinstimmt. Kriteriumsvalidität ist ein Gütekriterium von Tests und ein spezieller Aspekt der Konstruktvalidität.
- Die interne Validität schließlich bezieht sich auf die innere methodische Güte von Studien und die externe Validität auf deren äußere Generalisierbarkeit. Es geht hier also um Validität auf der Ebene ganzer Untersuchungen.

Zwischen den Parametern der Validität und der Reliabilität besteht ein wichtiger Zusammenhang, wonach eine hohe Reliabilität die Voraussetzung für eine hohe Validität darstellt (WPGS 2019).

Wie und auf welcher Basis kann man nun Aussagen und Vergleiche hinsichtlich der Genauigkeit intraoraler Röntgenaufnahmen zur Kariesdiagnostik machen? Hierfür muss man sich bestimmte Testparameter näher anschauen. Dabei gilt es, den Anteil an richtig-positiv/-negativ oder falsch-positiv/-negativ diagnostizierten kariösen Läsionen durch den Vergleich gegen einen Goldstandard, wie zum Beispiel die histologische Validation, zu ermitteln. Aus diesen Angaben wiederum lassen sich dann Sensitivität (Prozentsatz der als korrekt erkrankt erkannten Individuen) und Spezifität (Prozentsatz der als korrekt gesund erkannten Individuen) sowie der positive und negative Vorhersagewert (Anteil der Erkrankten innerhalb der Gruppe aller positiv bzw. negativ getesteten Individuen) errechnen (Abb. 1-2).

| | Krankheit: ja | Krankheit: nein | |
|---------------------|---|---|---|
| Test positiv | richtig-positiv | falsch-positiv | positiver Vorhersagewert: $\frac{\text{Anzahl richtig Positiver}}{\text{Anzahl aller positiv Getesteten}}$ |
| Test negativ | falsch-negativ | richtig-negativ | negativer Vorhersagewert: $\frac{\text{Anzahl richtig Negativer}}{\text{Anzahl aller negativ Getesteten}}$ |
| | Sensitivität $\frac{\text{Anzahl richtig Positiver}}{\text{Anzahl aller Kranken}}$ | Spezifität: $\frac{\text{Anzahl richtig Negativer}}{\text{Anzahl aller Gesunden}}$ | |

Abbildung 1-2: Die 4-Felder-Tafel (Meyer-Lückel et al., 2012).

Problematisch stellt sich in diesem Zusammenhang jedoch dar, dass bei der Durchführung einer Laborstudie zwar in allen Fällen eine histologische Validation erfolgen kann, bei der Durchführung einer klinischen Studie ist dies allerdings aus ethischen Gründen natürlich nicht möglich. Demnach werden in diesem Studien-Setting falsch-negative Testergebnisse, zum Beispiel weil Karies im Röntgenbild übersehen wurde, nicht berücksichtigt. Denn im Gegensatz zu In-vitro-Studien, bei denen der tatsächliche Zustand aller Flächen absolut gegen einen Goldstandard überprüft werden kann, ist es bei der Durchführung einer klinischen Studie aus ethischen Gründen nicht realisierbar, jedwede Zahnoberfläche aufzuziehen, um im Sinne einer korrekten Validation Aufschluss über den tatsächlichen Zustand der jeweiligen Zahnfläche zu erlangen.

Wie bereits weiter oben erwähnt, handelt es sich beim Röntgenverfahren um ein Diagnostikverfahren, welches schon seit sehr langer Zeit existiert und im Laufe der Zeit immer wieder optimiert und weiterentwickelt wurde. So wird es auch in Zukunft nicht aus der Zahnmedizin wegzudenken sein. Über die Jahre wurde eine große Anzahl an Studien veröffentlicht, welche die Genauigkeit dieses Diagnostikverfahrens zum Thema hatten. Jedoch fällt bei der Betrachtung jener Arbeiten auf, dass die Genauigkeitsangaben für ein und dasselbe Verfahren einer gewissen Heterogenität unterliegen. Möchte man diesbezüglich nun generelle Aussagen treffen, ist es daher zwingend notwendig, nicht nur einige wenige, sondern eine Vielzahl von Studien im Sinne einer Metaanalyse auszuwerten. Neben der Frage nach der Güte dieses althergebrachten Diagnostikverfahrens ist es ebenfalls hoch interessant, nach den Gründen für die Variation der hierfür angegebenen Werte zu forschen. Für die hier vorliegende Arbeit wurde daher im Sinne einer systematischen Metaanalyse und Übersichtsarbeit eine Vielzahl an Studien analysiert. Die einzelnen Veröffentlichungen wurden dabei nach bestimmten Informationen gescreent. Sensitivität und Spezifität spiegeln als Kennzahlen letztlich die Genauigkeit dieser Detektionsmethode wider. Zusätzlich zu diesen Kennzahlen mussten in den für diese Metaanalyse

einbezogenen Studien weitere wichtige Informationen enthalten sein. Hierauf soll im Folgenden näher eingegangen werden.

Die hier zugrundeliegenden Fragestellungen sollen letztendlich einerseits nach der Genauigkeit intraoraler Röntgenaufnahmen zur Kariesdiagnostik und andererseits vor allem nach der Identifikation von Ursachen für die große Variation der hierfür angegebenen Werte in den einzelnen Studien korrekt beantwortet werden können. Hierfür müssen die für die vorliegende Metaanalyse einbezogenen Studien zum einen Informationen enthalten, welche verlässlich sind, was als sogenannte interne Validität bezeichnet wird. Zum anderen müssen die gemachten Angaben natürlich auch klinisch anwendbar beziehungsweise übertragbar sein, dies wird die externe Validität genannt. Die interne Validität ist letztendlich eine Funktion von Bias und Präzision. Der Bias, also die Verzerrung oder der systematische Fehler, ist definiert als eine fortlaufende Abweichung einer beobachteten Messung von ihrem wahren Wert. Wenn in einer Studie ein Bias vorliegt, würde die betreffende Studie durchweg die wahren Parameter der Genauigkeit über- oder unterschätzen, wenn die Studie wiederholt werden würde. Demnach handelt es sich beim Bias um einen Fehler, welcher bei Wiederholung nicht ausgeglichen wird. Im Gegensatz hierzu handelt es sich bei der Präzision einer Messung um einen Zufallsfehler. Dieser gleicht sich bei Wiederholung aus. Sowohl ein Mangel an Präzision als auch das Auftreten eines Bias können eine Messung unverlässlich machen. Dies wiederum hat zur Folge, dass somit der wahre Sachverhalt des untersuchten Phänomens nicht korrekt wiedergespiegelt wird. Denn sowohl das Vorliegen eines Bias als auch ein Mangel an Präzision sind letztlich Quellen für Variationen. Folglich fehlt es einer Studie, in welcher ein Mangel an Präzision auftritt und zusätzlich ein Bias vorliegt, an interner Validität. Beide Parameter werden maßgeblich durch das Studiendesign beeinflusst. So wird der Zufallsfehler beispielsweise durch eine suffiziente Probengröße beziehungsweise einen ausreichend großen Patientenpool und einen soliden Versuchsaufbau bestimmt (Schmidt et al., 2013).

Der Bias, beziehungsweise das Verzerrungspotenzial, kann auf ganz verschiedenen Ebenen auftreten. Im Rahmen der in dieser Arbeit ausgewerteten „diagnostic accuracy studies“ kann dies zum Beispiel durch individuelle Testmessungen verursacht werden. Dieser Sachverhalt wird als sogenannter analytischer Bias bezeichnet. Anders formuliert, führen unterschiedliche Tests sozusagen zu unterschiedlichen Ergebnissen. Diese Erscheinung kann auch methodologische Ursachen haben, weshalb eine suffiziente Auswertung des Studiendesigns unumgänglich ist. Dies betrifft sowohl die interne Validität, welche sowohl die Bewertung der Gefahr des

Verzerrungspotentials als auch die Einschätzung des Grades der Variabilität einschließt, wie auch die externe Validität, also die Überprüfung der Anwendbarkeit der Studienergebnisse. Eine weitere Ursache dieser Erscheinung kann aber auch in methodologischen Problematiken liegen, welche letztlich durch das Studiendesign bestimmt werden. Eine suffiziente Auswertung der internen Validität schließt also demnach sowohl die Bewertung der Gefahr des Verzerrungspotenzials als auch die Einschätzung des Grades der Variabilität, welche durch den Zufallsfehler beeinflusst wird, mit ein.

Hierzu muss das Studiendesign beurteilt werden, was natürlich voraussetzt, dass diesbezüglich in der jeweiligen Studie präzise Angaben gemacht werden. Eine weitere wichtige Determinante zur Bewertung einer Studie ist schließlich die Anwendbarkeit ihrer Ergebnisse, also ihre sogenannte externe Validität.

Die Evaluation der Angaben von Sensitivität und Spezifität ermöglicht es dem Zahnarzt, das Risiko einer Über- beziehungsweise Unterbehandlung zu minimieren. Auch können auf Basis dieser Werte die verschiedenen zur Verfügung stehenden Detektionsmethoden besser verglichen werden. Weiterhin ist die Einschätzung jener Parameter aufgrund der in den einzelnen Altersgruppen jeweils unterschiedlichen Karieserfahrung relevant. Denn es besteht ein Zusammenhang zwischen dem Ausmaß der Karieserfahrung und dem Auftreten von Über- und Unterbehandlung (Schwendicke et al., 2014). Hierauf soll im folgenden Absatz basierend auf den Erkenntnissen der deutschen Mundgesundheitsstudie näher eingegangen werden.

Neben anderen wichtigen Erkrankungen der Mundhöhle wird die Karieserfahrung im Rahmen der deutschen Mundgesundheitsstudie in vier Alterskohorten untersucht. Innerhalb der Gruppe der 12-jährigen Kinder fällt dabei auf, dass sich die Zahl der kariesfreien Gebisse zwischen 1997 (DMS III) und 2014 (DMS V) verdoppelt hat. Von dieser positiven Entwicklung profitieren jedoch nicht sämtliche Angehörige aller sozialer Schichten gleichermaßen. Bei genauerer Betrachtung fällt somit auf, dass das Drittel mit der höchsten Karieserfahrung dreimal mehr erkrankte Zähne als das Drittel mit der niedrigsten Karieserfahrung aufweist. Man spricht in diesem Zusammenhang von der Polarisation der Karies beziehungsweise einer Ungleichheit der Karieserfahrung. Diese Entwicklung führt letztlich dazu, dass nur wenige Individuen besonders stark von kariösen Läsionen betroffen sind, wohingegen der größere Prozentsatz der untersuchten 12-jährigen Kinder kaum bis gar keine Läsionen aufweist (Marthaler et al., 1996; Marthaler TM, 2004). Leider kann es im Rahmen der Kariesdiagnostik naturgemäß immer wieder zu Fehldiagnosen kommen. Hierbei werden sogenannte Typ-1-Fehler (falsch-positives Testergebnis) und Typ-2-Fehler (falsch-

negatives Testergebnis) unterschieden. Während Typ-1-Fehler eine Überbehandlung zur Folge haben, führen Typ-2-Fehler letztlich zu einer Unterbehandlung. Somit wirkt sich die Polarisierung der Karies indirekt auch auf die Kariesdiagnostik aus. So führt ein Rückgang der Kariesprävalenz zum vermehrten Auftreten von Typ-1-Fehlern, während entsprechend bei einer Erhöhung vermehrt Typ-2-Fehler auftreten. Daher ist es wichtig, die verschiedenen zur Verfügung stehenden Detektionsmethoden anhand der Kennzahlen Sensitivität und Spezifität vergleichen zu können. Aufgrund der unterschiedlichen Karieserfahrung muss über die Anwendung verschiedener Detektionsmethoden beziehungsweise die Kombination unterschiedlicher Detektionsmethoden in Abhängigkeit von der jeweiligen Sensitivität und Spezifität nachgedacht werden.

Die Analyse der Studienlage hinsichtlich der Genauigkeit (Sensitivität und Spezifität) intraoraler Röntgenaufnahmen zur Kariesdiagnostik macht deutlich, dass hier eine große Variabilität zwischen den jeweils angegebenen Werten besteht (Bader et al., 2001). Das Ziel der hier vorliegenden Arbeit besteht demnach in einer systematischen Bewertung von Studien, welche Angaben zur Genauigkeit der radiologischen Kariesdiagnostik beinhalten, um diese schließlich einer Metaanalyse zu unterziehen. Um nun konkret mögliche Ursachen für die weiter oben erwähnte Variabilität der in den unterschiedlichen Studien angegebenen Genauigkeit intraoraler Röntgenaufnahmen zur Kariesdiagnostik zu untersuchen, sollen in der vorliegenden Arbeit verschiedene Parameter bewertet werden. Zum einen könnten für diesen Sachverhalt die klinischen Bedingungen

- Läsionstiefe,
- Dentition und
- jeweils betroffene Zahnfläche

oder aber auch methodologische Faktoren, wie

- beim Nachweistest zur Anwendung kommende Röntgenverfahren,
- zur Anwendung kommender Referenztest und
- Anzahl und Erfahrungsstand der Untersucher

verantwortlich sein. Das systematische Zusammenstellen der aus den diversen Studien verfügbaren Angaben zur Genauigkeit intraoraler Röntgenaufnahmen zur Kariesdiagnostik könnte Aufschluss über dessen Validität und auch den Einfluss möglicher Störfaktoren geben.

Hierbei soll eine separate Analyse für okklusale und proximale Läsionen erfolgen. Weiterhin soll der Einfluss der histologischen Läsionstiefe (initiale Läsion/jedwede Karies, Dentinläsion, kavitierte Läsion) untersucht werden. Letzteres ist insbesondere deshalb bedeutend, da sich hieraus die Notwendigkeit einer restaurativen Versorgung ergibt beziehungsweise diese Diagnose somit als Schwellenwert für bestimmte klinische Entscheidungen steht. Und schließlich soll auch beurteilt werden, ob die Dentition (Milchzähne/bleibende Zähne) sowie die Art des Röntgenverfahrens (analog/digital) Einfluss auf die Genauigkeit intraoraler Röntgenaufnahmen zur Kariesdiagnostik haben oder ob diesbezüglich bereits innerhalb einzelner Studien (das heißt zwischen den einzelnen Untersuchern oder den angewendeten Röntgenverfahren) große Unterschiede zwischen den jeweils angegebenen Werten bestehen.

2 Material und Methoden

Die in der vorliegenden Arbeit „Sensitivität und Spezifität intraoraler Röntgenaufnahmen zur Detektion von Karies“ zugrundeliegende Kernfrage nach der Genauigkeit dieses Diagnostikverfahrens ist entsprechend dem PICO-Modell, einem Hilfsschema aus der Evidenzbasierten Medizin, formuliert worden (Moher et al., 2009). PICO steht als Akronym für „Population“, „Intervention“, „Comparison“ und „Outcome“. Bezogen auf das in dieser Arbeit untersuchte Thema stehen diese vier Parameter im Falle von

- „Population“ für die untersuchte Studienpopulation beziehungsweise das jeweilige Studiensetting (klinische Studie oder Laborstudie).
- „Intervention“ für den angewendeten Indextest beziehungsweise intraorale Röntgenaufnahmen.
- „Comparison“ für den angewendeten Referenztest beziehungsweise die jeweilige Validationsmethode.
- „Outcome“ für das jeweilige Untersuchungsergebnis, also Karies ja/nein.

2.1 Einschlusskriterien

Es wurden sowohl klinische als auch In-vitro-Studien eingeschlossen, die die Detektion von Primärkaries mittels intraoraler röntgenologischer Verfahren zum Ziel hatten. Aufbau und Struktur dieser Übersichtsarbeit sind sowohl am „Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy“, als auch an den PRISMA-Richtlinien „Preferred reporting items for systematic reviews and meta-analyses“ orientiert. Folgende Einschlusskriterien wurden eingesetzt.

2.1.1 Population

Die eingeschlossenen Studien untersuchten im Falle klinischer Studien Patienten mittels Röntgenverfahren auf Primärkaries. In-vitro-Studien untersuchten menschliche Zähne mit

natürlichen kariösen Läsionen. In beiden Fällen wurden sowohl Milchzähne als auch bleibende Zähne einbezogen. Studien, welche sich mit der Untersuchung sekundärer oder künstlich erzeugter kariöser Läsionen befassten, wurden ausgeschlossen.

Bei der röntgenologischen Detektion kariöser Läsionen wird die Behandlungsentscheidung im klinischen Alltag von verschiedenen Faktoren abhängig gemacht. Zum einen spielt die zu beurteilende Fläche eine wichtige Rolle. Es wird hier zwischen Okklusal- und Approximalflächen differenziert. Während im Fall von Okklusalflächen aufgrund von Überlagerungseffekten, bedingt durch das Höcker-Fissuren-Relief, initiale Läsionen röntgenologisch schwerer zu detektieren sind, als bis ins mittlere Dentin reichende Defekte, ist dies bei Approximalflächen in der Regel besser möglich. In diesem Sinne mussten die Untersuchungsergebnisse der inkludierten Studien sowohl für Approximal- als auch Okklusalflächen separat zu entnehmen sein. Weiterhin dient als Schwellenwert für die Ableitung klinischer Behandlungsentscheidungen in der Regel die Läsionstiefe der kariösen Defekte. Während initiale Läsionen non- oder mikro-invasiv behandelt werden, sind bei kavitierten Läsionen restaurative Therapieformen das Mittel der Wahl. So ist ab einer röntgenologischen Ausdehnung der Läsion bis in das mittlere Dentindrittel in mehr als 50 Prozent der Fälle von einer Oberflächenkavität auszugehen (Meyer-Lückel et al., 2012). Als Schwellenwerte für die Analyse der einbezogenen Studien wurden für Approximal- und Okklusalflächen daher separat jeweils drei Parameter gewählt:

- jegliche Art von Karies ja/nein
- Dentinkaries ja/nein
- kavitierte Karies ja/nein

Die jeweils in den Studienpopulationen vorliegende Kariesprävalenz wurde nicht als Ausschlusskriterium verwendet. Die Auswertung der Ergebnisse der klinischen Studien und der In-vitro-Studien erfolgte aufgrund des grundsätzlichen Unterschiedes zwischen beiden Studiensettings jeweils separat.

2.1.2 Indextest

Als Diagnostikverfahren kamen intraorale Aufnahmeverfahren zum Einsatz. Dies schloss sowohl Bissflügelröntgenaufnahmen als auch periapikale Röntgenaufnahmen ein, da bei einer Vielzahl von Studien nicht klar zwischen den einzelnen Aufnahmemodalitäten unterschieden wurde.

2.1.3 Referenztest

Um gemäß der gewählten Schwellenwerte eine Einschätzung des tatsächlichen Zustandes der untersuchten Flächen zu erlangen, mussten den inkludierten Studien Aussagen über die jeweils gewählte Validationsmethode zu entnehmen sein. Hierfür wurden als Goldstandard destruktive und nicht destruktive Methoden unterschieden. Als destruktive Maßnahmen kamen die histologische Diagnostik, die Mikroradiografie und im Falle klinischer Studien die invasive Diagnostik zum Einsatz. Letztere Variante birgt allerdings die Gefahr, falsch-negative Ergebnisse auszublenden, wobei es sich hierbei um ein grundsätzliches ethisches Problem klinischer Studien handelt. Im Gegensatz zu In-vitro-Studien, wo jede Zahnoberfläche histologisch beziehungsweise mikroradiografisch untersucht werden kann, ist es im Falle klinischer Studien bis auf ganz bestimmte Ausnahmen aus ethischen Gründen nicht möglich, jede Zahnoberfläche zu validieren.

Als nicht destruktive Validationsmethode kam schließlich die visuell-taktile Diagnostik zur Anwendung, im Fall von Okklusalfächen ohne und bei Approximalfächen mit temporärer Separation. Bei letzterer Validationsmethode wurden kieferorthopädische Gummis oder Holzkeile eingesetzt, um Zähne temporär zu separieren. Somit konnten auch approximale Zahnoberflächen mit Spiegel und Sonde visuell-taktil überprüft werden und dadurch im klinischen Setting falsch-negative Ergebnisse ausgeschlossen werden.

2.1.4 Testergebnisse

Den eingeschlossenen Studien mussten die Angaben der richtig-positiven, falsch-positiven, richtig-negativen sowie der falsch-negativen Testergebnisse zu entnehmen sein. Konnten diese Angaben nicht direkt abgeleitet werden, wurden sie indirekt über die Angaben der Prävalenz, der Anzahl der untersuchten Flächen sowie der Sensitivität und der Spezifität vermittels einer 4-Felder-Tafel rekonstruiert. Studien, in denen lediglich die Auswertung von ROC-Kurven erfolgte, wurden nicht eingeschlossen.

2.2 Informationsquellen

Im Rahmen des Screenings wurden Manuskripte, welche bis September 2014 veröffentlicht waren, mittels Medline via PubMed und Embase via Ovid herausgesucht. Hierbei gab es hinsichtlich des Veröffentlichungsdatums und der Sprache des veröffentlichten Artikels keinerlei Einschränkungen.

Auch wurden über die in den einzelnen Studien angegebenen Referenzen per Handsuche Abhandlungen hinzugefügt, welche nicht über PubMed oder Embase verfügbar waren. Systematische Übersichtsarbeiten wurden über die Medion database gesucht, um von hier weitere Studien zu identifizieren. Graue Literatur wurde über opengrey.eu ergänzt (Deville et al., 2000).

2.3 Suche und Auswahl der Studien

Die Suchstrategie der hier vorliegenden Arbeit basierte auf den Empfehlungen der Cochrane Collaboration und setzte sich zusammen aus dem Indextest und bestimmten Zielparametern (Deville et al., 2000). Letztere beinhalteten sowohl die klinische Situation der Patientenauswahl als auch die für diese Arbeit benötigten Kennzahlen, also Sensitivität und Spezifität, als Schlüsselwörter. Um sicherzustellen, dass die Suche möglichst erfolgreich die Auswahl relevanter Studien begünstigte, wurden die gewünschten Begrifflichkeiten beziehungsweise deren Synonyme mittels der Booleschen Operatoren UND/ODER untereinander verknüpft. Die Suchstrategie wurde in Anlehnung an eine vergleichbare Veröffentlichung erstellt (Gimenez T, 2013). Im Detail setzte sich diese aus den gewünschten Zielparametern, der Methode der Kariesdetektion und schließlich der klinischen Situation zusammen:

- Zielparameter: detection OR roc OR sensitivity OR specificity OR predictive value OR receiver
- Klinische Situation: caries OR carious OR decay
- Methode der Kariesdetektion: bitewing OR radiograph OR radiography OR xray OR x-ray OR roentgen OR radiology OR radiologic OR radiographic

Anschließend wurden alle Titel und Abstracts durch zwei voneinander unabhängige Rezensenten hinsichtlich der Inklusionskriterien untersucht. Die somit eingegrenzte Auswahl an Veröffentlichungen wurde dann im Volltext kontrolliert. Ein- bzw. Ausschluss wurde durch beide Rezensenten diskutiert, gegebenenfalls wurde ein dritter Rezensent hinzugezogen.

2.4 Datenextraktion

Die Daten aus den letztlich inkludierten Studien wurden durch einen Rezensenten unter Verwendung von Excel-Tabellen extrahiert und gesammelt. Anschließend wurde die

Datenextraktion stichprobenartig bei 25 % der Studien (30 Veröffentlichungen) durch einen zweiten Rezensenten überprüft. Hierbei traten keinerlei Unstimmigkeiten auf.

Gemäß den Richtlinien der Cochrane Collaboration (Higgins et al., 2011) wurden den eingeschlossenen Veröffentlichungen folgende Angaben entnommen: Studie (Jahr der Veröffentlichung, Land, Studientyp – klinisch oder in vitro), Studienteilnehmer, untersuchte Zähne (Milchzähne, bleibende Zähne), Ort der Läsion (okklusal oder approximal), Nachweistest/Indextest (analoges/digitales Röntgen oder beides/unklar), Bezugstest/Referenztest (visuell-taktil oder histologisch/Mikroradiografie/Exkavation), Schweregrad der Läsion gemäß Bezugstest/Referenztest (jedwede Art von Läsion [sowohl Schmelz- als auch Dentinläsionen, da viele Studien hier nicht differenziert haben], Dentinläsionen, kavitierte Läsionen) und Ergebnisse der Studien (Prävalenz, Genauigkeit, Sensitivität, Spezifität, richtig-/falsch-negative/-positive Testergebnisse).

2.5 Verzerrungspotenzial

Das Verzerrungspotenzial wurde mithilfe der QUADAS-2-Richtlinien (Whiting et al., 2011) ermittelt. Gemäß dieser Richtlinien besteht eine dichotome Unterteilung in „Verzerrungspotenzial“ und „Bedenken hinsichtlich der Anwendbarkeit der Ergebnisse“ der einzelnen Studien. Bezogen auf das Verzerrungspotenzial findet eine weitere Unterteilung in vier Schlüssel-domänen statt. Die vier Schlüssel-domänen sind Patientenauswahl, Nachweistest, Bezugstest und zeitlicher Testverlauf. Jede Schlüssel-domäne besteht wiederum aus Signalfragen, die auf typische Gründe für einen Bias hinweisen sollen. Hierbei wurden folgende Aspekte berücksichtigt: Patientenauswahl (keine unbegründeten Ausschlüsse, keine Fall-Kontroll-Studie, zufälliger oder fortlaufender Einschluss von Patienten), Nachweistest (Verblindung gegenüber dem Referenztest, definierte Schwellenwerte), Bezugstest (valider Bezugstest, Auswertung unabhängig vom Nachweistest), zeitlicher Testverlauf (ausreichend Zeit zwischen Nachweis- und Bezugstest, alle Läsionen auf die entsprechende Referenz bezogen und alle Ergebnisse in die Auswertung einbezogen). Die Patientenauswahl (Patienten/Läsionen entsprechen der zugrundeliegenden Fragestellung, realistische Kariesprävalenz und Läsionsverteilung), der

Nachweistest (Test, Durchführung und Auswertung entsprechen der zugrundeliegenden Fragestellung) und der Bezugstest (diagnostizierter Zustand entspricht der zugrundeliegenden Fragestellung) wurden weiterhin hinsichtlich ihrer Anwendbarkeit auf die hier thematisierte Fragestellung untersucht. Studien, welche in allen Aspekten ein niedriges Verzerrungspotenzial aufwiesen wurden als „niedriges Risiko“ eingestuft, alle anderen als „hohes“ oder „unklares Risiko“ (Tabelle 2-1).

Tabelle 2-1: QUADAS-2-Richtlinien: Verzerrungspotenzial und Bedenken hinsichtlich der Anwendbarkeit der Ergebnisse, 1 = niedriges Risiko, 2 = hohes Risiko, 3 = unklares Risiko (Whiting et al., 2011).

| Studie | Verzerrungspotential | | | | Bedenken hinsichtlich der Anwendbarkeit | | |
|--------|----------------------|--------------|------------|------------------------|---|--------------|--------------|
| | Patientenauswahl | Nachweistest | Bezugstest | Zeitlicher Testverlauf | Patientenauswahl | Nachweistest | Referenztest |
| 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 |
| 2 | 1 | 1 | 1 | 1 | 2 | 1 | 1 |
| 3 | 2 | 3 | 1 | 1 | 2 | 1 | 1 |
| 4 | 2 | 3 | 1 | 1 | 2 | 1 | 1 |
| 5 | 2 | 3 | 1 | 1 | 2 | 1 | 1 |
| 6 | 2 | 3 | 1 | 1 | 2 | 3 | 1 |
| 7 | 2 | 3 | 1 | 1 | 2 | 1 | 1 |
| 8 | 2 | 3 | 1 | 1 | 2 | 3 | 1 |
| 9 | 2 | 3 | 1 | 2 | 2 | 1 | 1 |
| 10 | 1 | 3 | 1 | 2 | 1 | 1 | 1 |

2.6 Synthese

Als statistisches Merkmal diente in der hier vorliegenden Arbeit die jeweilige Zahnoberfläche, also die Okklusal- oder Approximalfläche. Zur Auswertung wurden die Sensitivität, die Spezifität und das Quotenverhältnis (diagnostic odds ratio oder DOR) aus dem primären Datensatz errechnet (über die gemittelte Genauigkeit). Beim sogenannten Quotenverhältnis handelt es sich um ein Maß dafür, um wie viel größer die Chance eines positiven Testergebnisses innerhalb einer Gruppe von Erkrankten ist, verglichen mit der Chance eines positiven Testergebnisses innerhalb einer Gruppe von nicht Erkrankten (Glas et al., 2003). Das Quotenverhältnis kann Werte zwischen 0 und unendlich annehmen, wobei 0 den schlechtesten diagnostischen Wert darstellt. Der Wert 1 spricht für ein gleiches Quotenverhältnis (zufällige Genauigkeit), Werte größer 1 sprechen für eine verwertbare Testgenauigkeit (Glas et al., 2003). Der Vorteil der Anwendung des Quotenverhältnisses besteht in der Unabhängigkeit von der Prävalenz und darin, dass Angaben zur Genauigkeit mit einer Größe erfasst werden.

Die Synthese der Daten erfolgte separat für verschiedene Zahnoberflächen (okklusal, approximal) und Schwellenwerte (jedwede Art von Läsion, Läsionen bis ins Dentin, kavitierte Läsionen). Zur Metaanalyse kam der DerSimonian-Laird-Schätzer zum Einsatz. Im Rahmen der Hauptanalyse (in jeder Studie angegebene gemittelte Genauigkeit) wurde bei der Auswertung weiterhin zwischen klinischen Studien und In-vitro-Studien unterschieden. Zur Darstellung der gesammelten Daten wurden Summary-receiver-operator-characteristic-Kurven (sROC) verwendet. Diese Art der Darstellung ermöglicht es, Austauschbeziehungen zwischen der Sensitivität und der Spezifität bei verschiedenen Schwellenwerten grafisch darzustellen. Durch das Integrieren der Fläche unterhalb der ROC-Kurve (AUC) wird schließlich die diagnostische Genauigkeit dargestellt. Dabei stellen Werte von 1 und 0 die jeweils beste beziehungsweise schlechteste Testgenauigkeit dar, während ein Wert von 0,5 für zufällige Testergebnisse spricht (Hanley et al., 1982). Weiterhin wurde der Schwellenwert berechnet, bei dem die Sensitivität und die Spezifität den gleichen Wert annehmen (Q^*) (Walter SD, 2002). Als Fehlerschätzer für DOR und sROC kamen die 95% Konfidenzintervalle (95 % KI) zur Anwendung, während für AUC und Q^* eine Standardabweichung (SE) angewendet wurde.

Als Heterogenitätsmaße zur Abschätzung der Variabilität innerhalb der ausgewerteten Studien wurden Cochran's-Q- und I^2 -Teststatistiken angewendet (Higgins et al., 2002), wobei die Heterogenität über DORs erfasst wurde. Funnel-plot-Analysen und einseitige Egger-Regressionstests wurden verwendet, um den Einfluss kleiner Studien oder eines Publikationsbias zu ermitteln (Higgins et al., 2011; Egger et al., 1997). Metaregressionsanalysen wurden durchgeführt, um den Einfluss möglicher Störgrößen (Dentition, Jahr der Veröffentlichung, Referenztest, Publikationsbias/Anwendbarkeit) auf die DORs zu ermitteln (Duval S & Tweedie R, 2000). Dazu kam die Unrestricted-Maximum-Likelihood-Methode zum Einsatz. Es kamen sowohl uni- als auch multivariate Regressionsanalysen zur Anwendung. Zur Synthese der gesammelten Daten kamen MetaDisc 1.4 (Unidad de Bioestadística Clínica del Hospital Ramón y Cajal, Madrid, Spain) und Comprehensive Meta-Analysis 2.2.64 (Biostat, Englewood, USA) zum Einsatz.

3 Ergebnisse

3.1 Suchergebnisse

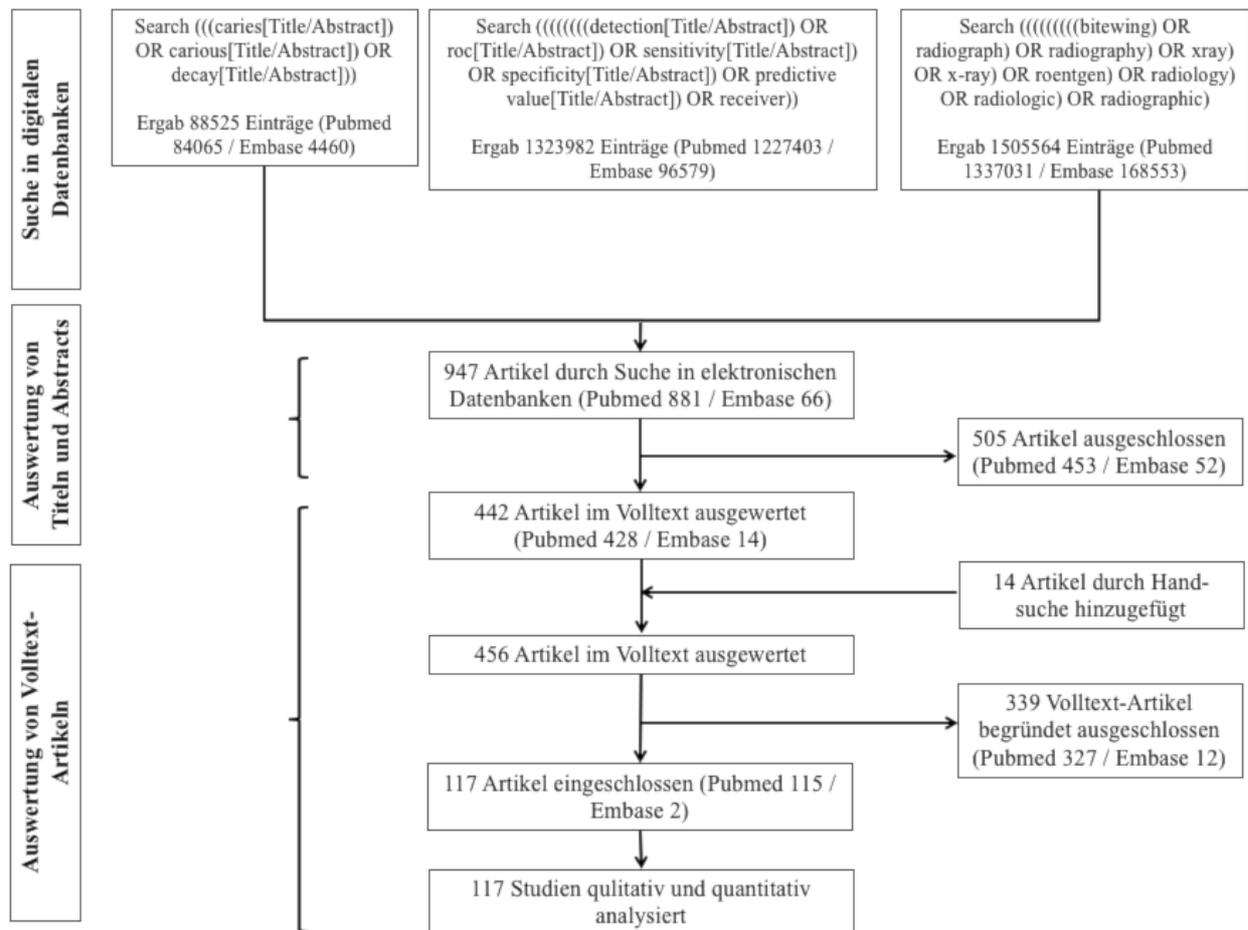


Abbildung 3-1: Der Suchprozess (Schwendicke et al., 2015).

Über die Anwendung elektronischer Datenbanken wurden 947 Studien als potenziell verwendbar erfasst. Schlussendlich wurden dann 442 Studien im Volltext analysiert. Durch Handsuche und Querverweise wurden im Nachgang 14 weitere Artikel hinzugefügt, sodass letztlich 456 Artikel im Volltext ausgewertet werden konnten. Dies führte zum Ausschluss von 339 Studien. Eine genaue Aufschlüsselung über die Hintergründe hierfür findet sich im Anhang. Innerhalb der grauen Literatur konnten keinerlei relevante Studien erfasst werden. Letztlich wurden 117 Studien im Rahmen der Metaanalyse einbezogen und ausgewertet (Abb. 3-1 und Tabelle 3-1).

Tabelle 3-1: Eingeschlossene Studien nach QUADAS-2-Richtlinienn (grün hinterlegt sind Studien mit ausschließlich niedrigem Risiko) : Verzerrungspotenzial und Bedenken hinsichtlich der Anwendbarkeit der Ergebnisse, 1 = niedriges Risiko, 2 = hohes Risiko, 3 = unklares Risiko (in Anlehnung an Whiting et al., 2011)

| Studie | Verzerrungspotenzial | | | | Bedenken hinsichtlich der Anwendbarkeit | | | |
|-------------------------------------|----------------------|------------------|--------------|------------|---|------------------|--------------|--------------|
| | Pubmed | Patientenauswahl | Nachweistest | Bezugstest | Zeitlicher Testverlauf | Patientenauswahl | Nachweistest | Referenztest |
| 1. Sansare K et al., 2014 | | 1 | 1 | 1 | 3 | 1 | 1 | 1 |
| 4. Sinanoglu A et al., 2014 | | 1 | 1 | 3 | 1 | 1 | 1 | 3 |
| 46. Çınar et al., 2013 | | 1 | 1 | 3 | 1 | 1 | 1 | 3 |
| 51. Shimada Y et al., 2013 | | 1 | 1 | 3 | 3 | 1 | 1 | 3 |
| 55. Abesi F et al., 2012 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 65. Ástvaldsdóttir et al., 2012 | | 1 | 1 | 3 | 1 | 1 | 1 | 3 |
| 78. Jablonski-Momeni A et al., 2012 | | 3 | 1 | 3 | 3 | 3 | 1 | 3 |
| 79. Pakkala T et al., 2012 | | 1 | 1 | 3 | 3 | 1 | 1 | 3 |
| 81. Mendes FM et al., 2012 | | 3 | 1 | 3 | 3 | 3 | 1 | 3 |
| 85. Kamburoglu K et al., 2012 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 87. Novaes TF et al., 2012 | | 2 | 1 | 1 | 3 | 2 | 1 | 1 |
| 88. Souza JF et al., 2012 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 100. Diniz MB et al., 2012 | | 1 | 1 | 1 | 3 | 1 | 1 | 1 |
| 104. Graye M et al., 2012 | | 1 | 1 | 3 | 3 | 1 | 1 | 3 |
| 117. Novaes TF et al., 2012 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 123. Rathore S et al., 2012 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 882. Hintze H1 et al., 1998 | | 2 | 2 | 3 | 3 | 2 | 2 | 3 |
| 136. Chen J et al., 2011 | | 1 | 1 | 3 | 3 | 1 | 1 | 3 |
| 139. Soviero VM et al., 2011 | | 1 | 1 | 3 | 1 | 1 | 1 | 3 |
| 141. Tracy KD et al., 2011 | | 1 | 1 | 1 | 3 | 1 | 1 | 1 |
| 142. Behere RR et al., 2011 | | 1 | 1 | 1 | 3 | 1 | 1 | 1 |
| 144. Xavier CR et al., 2011 | | 1 | 1 | 3 | 3 | 1 | 1 | 3 |
| 149. Diniz MB et al., 2011 | | 3 | 1 | 1 | 3 | 3 | 1 | 1 |

3 Ergebnisse

| Studie | Patientenauswahl | | | | Nachweistest | | | |
|--|------------------|------------------|--------------|------------|------------------------|------------------|--------------|--------------|
| | Pubmed | Patientenauswahl | Nachweistest | Bezugstest | Zeitlicher Testverlauf | Patientenauswahl | Nachweistest | Referenztest |
| 154. Bottenberg P et al., 2011 | | 1 | 1 | 3 | 1 | 1 | 1 | 3 |
| 155. Matos R et al., 2011 | | 3 | 1 | 1 | 1 | 3 | 1 | 1 |
| 163. Ekstrand KR et al., 2011 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 167. Kavvadia K et al., 2011 | | 1 | 1 | 3 | 1 | 1 | 1 | 3 |
| 173. Senel B et al., 2010 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 180. Novaes TF et al., 2010 | | 3 | 1 | 1 | 1 | 3 | 1 | 1 |
| 182. Holtzman JS et al., 2010 | | 1 | 1 | 1 | 3 | 1 | 1 | 1 |
| 183. Pontual AA et al., 2010 | | 1 | 1 | 3 | 3 | 1 | 1 | 3 |
| 189. Dias da Silva PR et al., 2010 | | 1 | 1 | 1 | 3 | 1 | 1 | 1 |
| 193. Mitropoulos P et al., 2010 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 197. Zangoeei Booshehry M et al., 2010 | | 1 | 1 | 3 | 1 | 1 | 1 | 3 |
| 202. Neuhaus KW et al., 2010 | | 1 | 1 | 3 | 1 | 1 | 1 | 3 |
| 216. Apostolopoulou D et al., 2009 | | 2 | 1 | 3 | 3 | 2 | 1 | 3 |
| 229. Isidor S et al., 2009 | | 1 | 1 | 1 | 3 | 1 | 1 | 1 |
| 234. Goel A et al., 2009 | | 1 | 1 | 3 | 3 | 1 | 1 | 3 |
| 241. Braga MM et al., 2009 | | 1 | 1 | 1 | 3 | 1 | 1 | 1 |
| 243. Kühnisch J et al., 2009 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 244. Young SM et al., 2009 | | 1 | 1 | 3 | 3 | 1 | 1 | 3 |
| 249. Diniz MB et al., 2009 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 265. Chu CH et al., 2009 | | 1 | 1 | 3 | 1 | 1 | 1 | 3 |
| 267. Rockenbach MI et al., 2008 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 274. Novaes TF et al., 2009 | | 1 | 1 | 1 | 3 | 1 | 1 | 1 |
| 276. Haiter-Neto F et al., 2009 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 277. Costa AM et al., 2008 | | 1 | 1 | 3 | 3 | 1 | 1 | 3 |
| 295. Schulze RK et al., 2008 | | 1 | 1 | 3 | 1 | 1 | 1 | 1 |

| Studie | Patientenauswahl | | | | Nachweistest | | | |
|------------------------------------|------------------|------------------|--------------|------------|------------------------|------------------|--------------|--------------|
| | Pubmed | Patientenauswahl | Nachweistest | Bezugstest | Zeitlicher Testverlauf | Patientenauswahl | Nachweistest | Referenztest |
| 300. Rodrigues JA et al., 2008 | | 1 | 1 | 3 | 3 | 1 | 1 | 3 |
| 304. da Silva Neto JM et al., 2008 | | 1 | 1 | 3 | 3 | 1 | 1 | 3 |
| 309. Kavvadia K et al., 2008 | | 1 | 1 | 1 | 3 | 1 | 1 | 1 |
| 316. Haiter-Neto F et al., 2008 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 319. Schulte AG et al., 2008 | | 1 | 1 | 3 | 1 | 1 | 1 | 3 |
| 325. Haiter-Neto F et al., 2007 | | 1 | 1 | 1 | 3 | 1 | 1 | 1 |
| 339. Berkhout WE et al., 2007 | | 1 | 1 | 3 | 3 | 1 | 1 | 3 |
| 346. Manton DJ et al., 2007 | | 1 | 1 | 3 | 3 | 1 | 1 | 3 |
| 356. Kalathingal SM et al., 2007 | | 1 | 1 | 3 | 1 | 1 | 1 | 3 |
| 358. Matalon S et al., 2007 | | 1 | 1 | 1 | 3 | 1 | 1 | 1 |
| 366. Kutcher MJ et al., 2006 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 369. Olmez A et al., 2006 | | 1 | 1 | 3 | 3 | 1 | 1 | 3 |
| 389. Mendes FM et al., 2006 | | 1 | 1 | 3 | 3 | 1 | 1 | 3 |
| 392. Virajsilp V et al., 2005 | | 1 | 1 | 3 | 3 | 1 | 1 | 3 |
| 396. Lussi A et al., 2006 | | 1 | 1 | 3 | 3 | 1 | 1 | 3 |
| 406. Burin C et al., 2005 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 413. Angnes V et al., 2005 | | 1 | 1 | 1 | 3 | 1 | 1 | 1 |
| 433. Mestriner SF et al., 2005 | | 1 | 1 | 3 | 1 | 1 | 1 | 3 |
| 434. Erten H et al., 2005 | | 1 | 1 | 3 | 3 | 1 | 1 | 3 |
| 435. Reis A et al., 2004 | | 1 | 1 | 1 | 3 | 1 | 1 | 1 |
| 458. Schulze RK et al., 2004 | | 1 | 1 | 3 | 2 | 1 | 1 | 3 |
| 462. Wojtowicz PA et al., 2003 | | 1 | 1 | 1 | 3 | 1 | 1 | 1 |
| 467. Rocha RO et al., 2003 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 469. Haak R et al., 2003 | | 1 | 1 | 3 | 3 | 1 | 1 | 3 |

| Studie | Verzerrungspotential | | | | Bedenken hinsichtlich der Anwendbarkeit | | | |
|---------------------------------------|----------------------|------------------|--------------|------------|---|------------------|--------------|--------------|
| | Pubmed | Patientenauswahl | Nachweistest | Bezugstest | Zeitlicher Testverlauf | Patientenauswahl | Nachweistest | Referenztest |
| 480. Matalon S et al., 2003 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 487. Costa AM et al., 2002 | | 1 | 1 | 3 | 1 | 1 | 1 | 3 |
| 490. Lussi A et al., 2003 | | 1 | 1 | 3 | 1 | 1 | 1 | 3 |
| 493. Heinrich-Weltzien R et al., 2002 | | 1 | 1 | 3 | 3 | 1 | 1 | 3 |
| 502. Wenzel A et al., 2002 | | 3 | 1 | 3 | 3 | 3 | 1 | 3 |
| 519. Mileman PA et al., 2002 | | 1 | 1 | 1 | 3 | 1 | 1 | 1 |
| 524. Espelid I et al., 2001 | | 1 | 1 | 1 | 3 | 1 | 1 | 1 |
| 527. Nair MK et al., 2001 | | 1 | 1 | 3 | 1 | 1 | 1 | 3 |
| 537. Haak R et al., 2001 | | 1 | 1 | 3 | 1 | 1 | 1 | 3 |
| 540. Attrill DC et al., 2001 | | 1 | 1 | 1 | 3 | 1 | 1 | 1 |
| 541. Lussi A et al., 2001 | | 1 | 1 | 1 | 3 | 1 | 1 | 1 |
| 557. Svanaes DB et al., 2000 | | 1 | 1 | 3 | 1 | 1 | 1 | 3 |
| 558. Côrtes DF et al., 2000 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 883. Duncan RC1 et al., 1995 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 602. Huysmans MC et al., 1998 | | 1 | 1 | 1 | 3 | 1 | 1 | 1 |
| 606. Skodje F et al., 1998 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 610. Firestone AR et al., 1998 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 611. Ashley PF et al., 1998 | | 1 | 1 | 1 | 3 | 1 | 1 | 1 |
| 637. Huysmans MC et al., 1997 | | 1 | 1 | 3 | 3 | 1 | 1 | 3 |
| 640. Ricketts DN et al., 1997 | | 1 | 1 | 1 | 3 | 1 | 1 | 1 |
| 643. White SC et al., 1997 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 645. Cayley AS et al., 1997 | | 1 | 1 | 1 | 3 | 1 | 1 | 1 |
| 649. Kay EJ et al., 1996 | | 3 | 1 | 1 | 3 | 3 | 1 | 1 |
| 653. Nielsen LL et al., 1996 | | 1 | 1 | 1 | 3 | 1 | 1 | 1 |

| Studie | Patientenauswahl | | | | Nachweistest | | |
|-------------------------------------|-------------------------|---------------------|-------------------|-------------------------------|-------------------------|---------------------|---------------------|
| Pubmed | Patientenauswahl | Nachweistest | Bezugstest | Zeitlicher Testverlauf | Patientenauswahl | Nachweistest | Referenztest |
| 664. Hintze H et al., 1996 | 3 | 1 | 3 | 3 | 3 | 1 | 3 |
| 693. Lazarchik DA et al., 1995 | 1 | 1 | 3 | 3 | 1 | 1 | 3 |
| 694. Ricketts DN et al., 1995 | 3 | 1 | 3 | 3 | 1 | 1 | 3 |
| 718. Espelid I et al., 1994 | 1 | 1 | 1 | 3 | 1 | 1 | 1 |
| 723. Verdonschot EH et al., 1993 | 1 | 1 | 3 | 3 | 1 | 1 | 3 |
| 730. Ketley CE et al., 1993 | 1 | 1 | 1 | 3 | 1 | 1 | 1 |
| 735. Russell M et al., 1993 | 1 | 1 | 1 | 3 | 1 | 1 | 1 |
| 737. Peers A et al., 1993 | 1 | 1 | 3 | 1 | 1 | 1 | 3 |
| 740. Lussi A et al., 1993 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 741. Wenzel A et al., 1992 | 1 | 1 | 3 | 3 | 1 | 1 | 3 |
| 746. Nyttun RB et al., 1992 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 747. Kay EJ et al., 1992 | 3 | 1 | 1 | 3 | 3 | 1 | 1 |
| 756. Verdonschot EH et al., 1992 | 3 | 1 | 3 | 3 | 3 | 1 | 3 |
| 765. Wenzel A et al., 1991 | 2 | 1 | 3 | 1 | 2 | 1 | 3 |
| 833. Espelid I et al., 1986 | 3 | 1 | 3 | 3 | 3 | 1 | 3 |
| 886. van Amerongen JP1 et al., 1992 | 1 | 1 | 1 | 3 | 1 | 1 | 1 |
| 893. Downer MC et al., 1975 | 2 | 1 | 3 | 3 | 2 | 1 | 3 |
| Studie | Patientenauswahl | | | | Nachweistest | | |
| Embase | Patientenauswahl | Nachweistest | Bezugstest | Zeitlicher Testverlauf | Patientenauswahl | Nachweistest | Referenztest |
| 45. Hibst R et al., 2001 | 2 | 1 | 3 | 3 | 2 | 1 | 3 |
| 49. Arijj Y et al., 1998 | 2 | 1 | 1 | 3 | 2 | 1 | 1 |

3.2 Eingeschlossene Studien und Verzerrungspotenzial

Die in der vorliegenden Arbeit eingeschlossenen Studien wurden im Zeitraum zwischen 1975 und 2014 erstellt. Insgesamt wurden dabei 13.375 Zähne und 19.108 Zahnflächen untersucht. Von den 117 inkludierten Studien wurden 24 in vivo und 93 in vitro durchgeführt. Hierbei wurden in den klinischen Studien insgesamt 1.755 Patienten untersucht. In 54 Studien wurde okklusale Karies und in 55 Studien proximale Karies untersucht, wobei 8 Studien beides zum Gegenstand hatten. 92 Studien hatten Karies in den bleibenden Zähnen, 23 in Milchzähnen und 2 Studien Karies in beiden Dentitionen zum Gegenstand. Als Referenztest kam in 106 Studien jeweils eine Art von destruktiver Validation zur Anwendung (histologische Validation, mikroradiografische Untersuchung oder Exkavation). In 11 Studien kam lediglich die visuell-taktile Validation zur Anwendung. 67 beziehungsweise 25 Studien erwähnten das analoge und digitale Röntgenverfahren, die verbleibenden Studien nannten entweder beides oder gaben nicht eindeutig an, welches System zur Anwendung kam.

Das Verzerrungspotenzial beziehungsweise die eingeschränkte Anwendbarkeit bezüglich der zugrundeliegenden Fragestellung wurde bei 23 Studien für niedrig und im Falle von 94 als hoch eingeschätzt. Am häufigsten kam letztere Einschätzung durch den Aspekt flow/timing zustande. In den meisten Studien erfolgten Referenz- und Bezugstest in zu kurzem Abstand oder aber es wurden diesbezüglich keinerlei Angaben gemacht (59 %). Weitere Ursachen für ein erhöhtes Verzerrungspotenzial begründeten sich durch den jeweils angewendeten Referenztest oder in der Auswahl der Patienten, was wiederum häufig Einfluss auf die jeweils vorliegende Kariesprävalenz beziehungsweise das Spektrum der vorliegenden Läsionen genommen hat (Tabelle 3-1).

3.3 Datenauswertung

Im Rahmen der Hauptanalyse (gemittelte accuracy) lag die zusammengefasste Sensitivität zur Erfassung jeglicher Art von okklusaler Karies bei 0,35 (0,95 % CI: 0,31/0,40) beziehungsweise 0,41 (0,39/0,44) für klinische Studien und In-vitro-Studien. Die zusammengefasste Spezifität lag bei 0,78 (0,73/0,83) und 0,70 (0,76/0,84). Für die Erfassung jeglicher Art kariöser Läsionen im approximalen Bereich lag die Sensitivität bei klinischen Studien beziehungsweise In-vitro-Studien

bei 0,24 (0,21/0,26) und 0,42 (0,88/0,90). Die diagnostic odds ratios (DOR) überschritten alle signifikant den Wert 1, was für eine suffiziente Testgenauigkeit spricht, und waren bei approximalen Läsionen größer als bei okklusalen Läsionen. Die Heterogenität war moderat ($I^2 > 50-67\%$ [50–60 = moderat; 50–90 = erhebliche Heterogenität]). Die Fläche unter der ROC-Kurve (AUC) bewegte sich zwischen 0,66 und 0,79 und die Schwelle, bei welcher Sensitivität und Spezifität gleich waren (Q^*), lag zwischen 0,62 und 0,73. Für diese Analyse wurden insgesamt zwischen 4 und 40 Studien herangezogen (Tab. 3-2).

Tabelle 3-2: Sensitivität, Spezifität und Genauigkeit röntgenologischer Kariesdetektion für die verschiedenen Schwellenwerte und Zahnflächen (Schwendicke et al., 2015).

| Läsionsart | Zahnfläche | Klinische Studien | | | | | In vitro-Studien | | | | |
|---------------------------|------------|--------------------|---------------------|---------------------|---------------------|--------------------------------------|--------------------|---------------------|---------------------|---------------------|--|
| | | Anzahl der Studien | Sensitivität | Spezifität | DOR | Heterogenität | Anzahl der Studien | Sensitivität | Spezifität | DOR | Heterogenität |
| Jegliche Art von Läsionen | Okklusal | 5 | 0.35 (0.31/0.40) | 0.78 (0.73/0.83) | 2.6 (0.80/8.2) | $I^2=67\%$; $Q=13$ ($p<0.05$) | 19 | 0.41 (0.39/0.44) | 0.80 (0.76/0.84) | 2.4 (1.4/3.9) | $I^2=50\%$; $Q=36$ ($p<0.01$) |
| | Approximal | 4 | 0.24 (0.21/0.26) | 0.97 (0.95/0.98) | 11.3 (3.9/32) | $I^2=56\%$; $Q=6.8$ ($p>0.05$) | 40 | 0.43 (0.41/0.45) | 0.89 (0.88/0.90) | 6.1 (4.6/8.22) | $I^2=66\%$; $Q=113$ ($p<0.001$) |
| Dentinläsionen | Okklusal | 15 | 0.56 (0.52/0.59) | 0.95 (0.94/0.96) | 26,1 (12,0/56,7) | $I^2=76\%$; $Q=59$ ($p<0.001$) | 45 | 0.56 (0.53/0.59) | 0.87 (0.85/0.89) | 8.3 (6.1/11.3) | $I^2=58\%$; $Q=105$ ($p<0.001$) |
| | Approximal | 2 | 0.36 (0.24/0.49) | 0.94 (0.89/0.97) | 7.5 (3.4/16.5) | $I^2=0\%$; $Q=0.5$ ($p>0.05$) | 24 | 0.45 (0.41/0.48) | 0.95 (0.94/0.96) | 16.0 (11.5/22.4) | $I^2=44\%$; $Q=41$ ($p<0.05$) |
| Kavitierte Läsionen | Okklusal | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| | Approximal | 6 | 0.64 (0.59/0.70) | 0.98 (0.97/0.99) | 56.3 (14.0/225) | $I^2=91\%$; $Q=57$ ($p<0.001$) | 6 | 0.61 (0.53/0.68) | 0.90 (0.87/0.93) | 15.2 (4.81/48.0) | $I^2=81\%$; $Q=26.2$ ($p<0.001$) |

Bei der Auswertung der Genauigkeit für die Detektion von Dentinläsionen erhöhten sich die ermittelten Sensitivitätswerte gegenüber der vorgenannten Analyse und reichten von 0,36 (0,24/0,49) für approximale Läsionen bis 0,56 (0,53/0,59) für okklusale Läsionen. Die Spezifität betrug zwischen 0,87 (0,85/0,89) und 0,95 (0,94/0,96) für die Detektion von Dentinläsionen (Tab. 3-1). Die DOR-Werte waren hoch (bis zu 26,1 [12,0/56,7] bei der Detektion von Dentinläsionen unter klinischen Bedingungen), während sich die Heterogenität sehr variabel zeigte ($I^2 = 0-76\%$). Für alle Analysen, abgesehen von der Detektion approximaler Läsionen unter In-vitro-

Bedingungen (2 Studien), ergaben sich AUC-Werte zwischen 0,81 und 0,93. Die Werte für Q^* lagen zwischen 0,74 und 0,77. Insgesamt wurden für diese Analyse 2 bis 45 Studien herangezogen (Tab. 3-1).

Hinsichtlich der Detektion kavierter okklusaler Läsionen konnten keine Studien ausgemacht werden. Bezogen auf approximale Läsionen wurden für diesen Schwellenwert in allen Studien, bis auf eine Ausnahme (Sansare et al., 2014), zusätzlich zur Genauigkeit hinsichtlich der Detektion jener Läsionen auch Angaben zur Korrektheit der Entscheidung kavitiert ja/nein gemacht. In allen Studien mit diesem Schwellenwert galt eine röntgenologische Dentinbeteiligung der Läsion als Schwellenwert für diese Diagnoseentscheidung. Die Sensitivität bei der Diagnose kavierter Läsionen war höher als bei der Detektion von jeglicher Karies beziehungsweise von Dentinläsionen. Die Werte lagen hier über 0,60. Die Spezifität lag weiterhin hoch (Werte über 0,90). Die DOR-Werte waren ebenfalls weiterhin hoch (15,2 [4,8/48] und 56,3 [14,0/225] unter In-vitro- beziehungsweise In-vivo-Bedingungen). Die Heterogenität war hoch ($I^2 > 81\%$). Die AUC-Werte bewegten sich zwischen 0,93 und 0,96 und Q^* lag zwischen 0,86 bis 0,91. Sechs Studien wurden für jene beiden Analysen herangezogen (Tab. 3-1). Bei allen Analysen lagen die Werte für die gesammelten Sensitivitäten unterhalb der Werte der gesammelten Spezifitäten (Abb. 3-2).

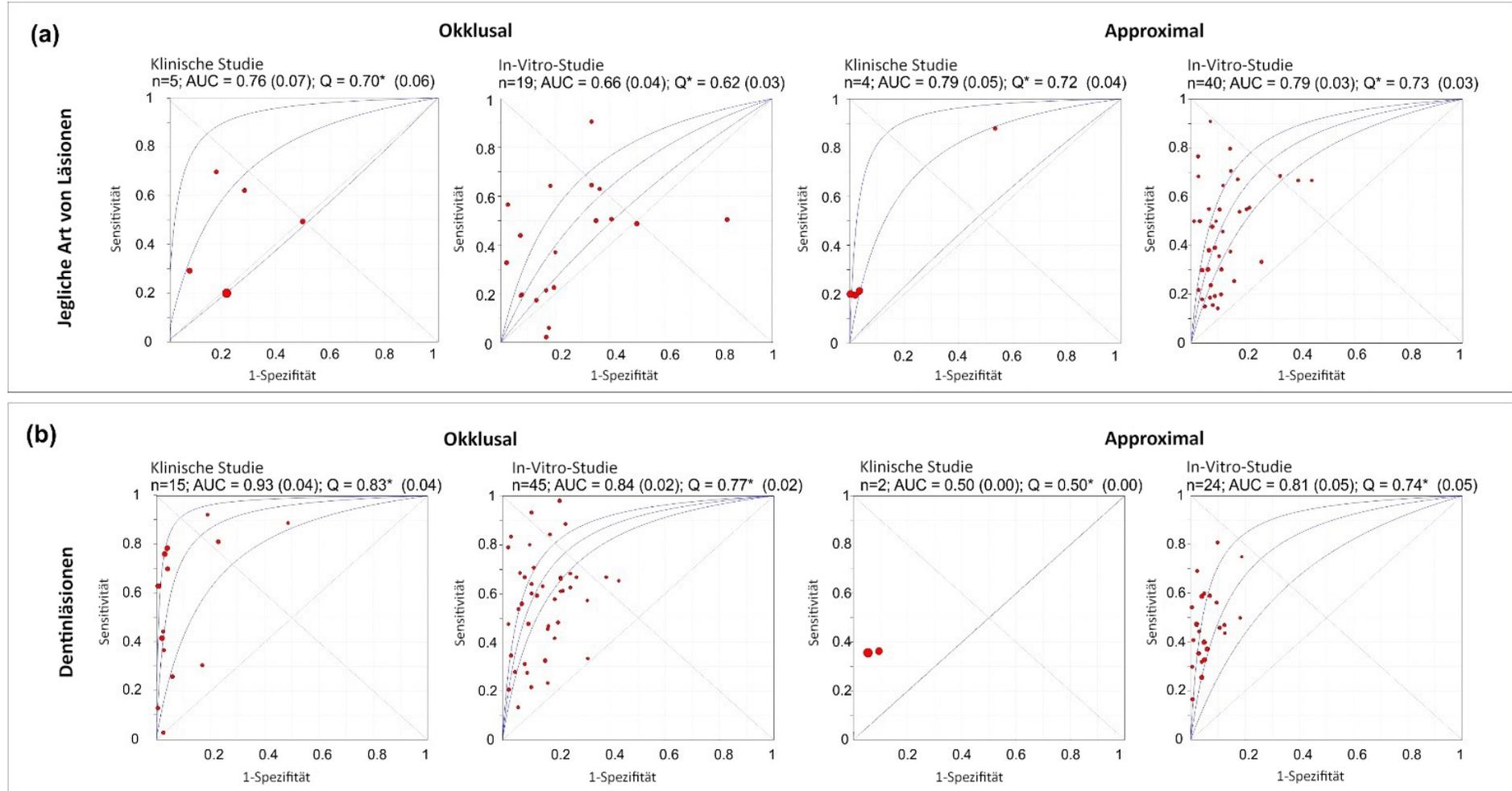


Abbildung 3-2.1: Erläuterung auf der folgenden Seite.

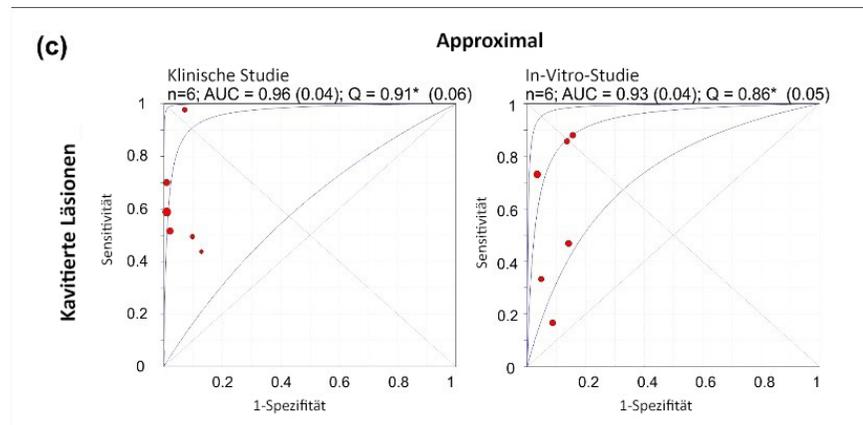


Abbildung 3-2.2: ROC-Kurven (Receiver Operating Characteristics). Es wurden Symmetrische ROCs wurden aufgezeichnet. Die drei Diagramme zeigen gepoolte ROC- und Konvidenzintervalle. Sensitivität und 1-Spezifität wurden gegeneinander aufgetragen. Der Bereich unter der Kurve AUC (in Klammern: SE) und der Schwellenwert, bei dem Sensitivität und Spezifität gleich sind (Q^*), werden hier dargestellt. AUC-Werte von 1 und 0 bedeuten die bestmögliche und schlechteste Genauigkeit, Werte von 0,5 bedeuten zufällige Genauigkeit. Es wurden getrennt Analysen für (a) jegliche Art von Läsionen, (b) Dentinläsionen und (c) kavitierte Läsionen durchgeführt. Dabei wurden separate Analysen für okklusale und proximale Zahnflächen durchgeführt und es wurde zwischen klinischen und Laborstudien unterschieden. (Schwendicke et al., 2015).

3.4 Publikationsbias

Die Gefahr eines Publikationsbias wurde für die einzelnen zuvor beschriebenen Subgruppenanalysen mittels Funnel-Plots und Egger-Regressionstests ermittelt. Insbesondere unter In-vitro-Bedingungen angefertigte Studien wurden hierbei als besonders anfällig für einen Publikationsbias befunden. Das Zurechnen eventuell fehlender Studien hatte keinen signifikanten Einfluss auf die errechneten DOR-Werte. Die einzige Ausnahme hierfür stellte die Subgruppenanalyse für den Schwellenwert jeglicher Karies unter klinischen Bedingungen dar. Diese Feststellungen wurden durch den Egger-Regressionstest bestätigt (Abb. 3-3 a und b).

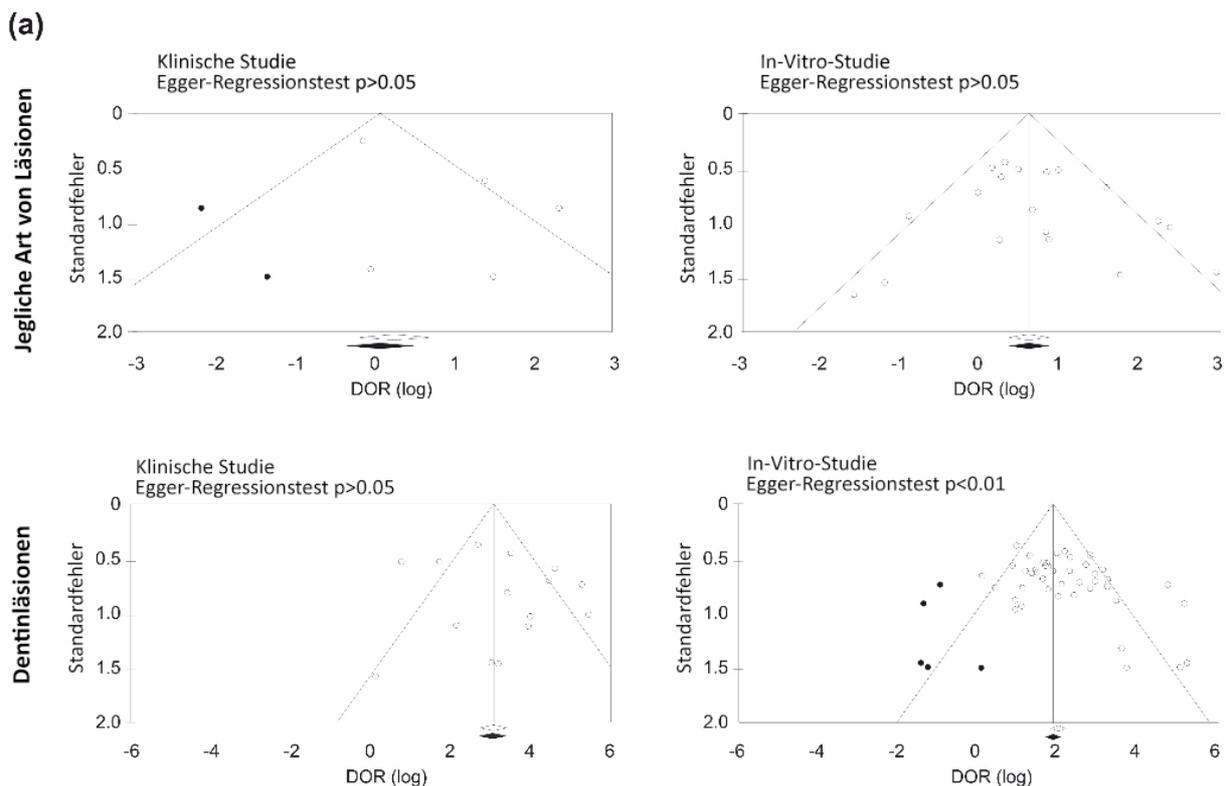


Abbildung 3-3 a: Funnel-Plots für okklusale Läsionen (Schwendicke et al., 2015).

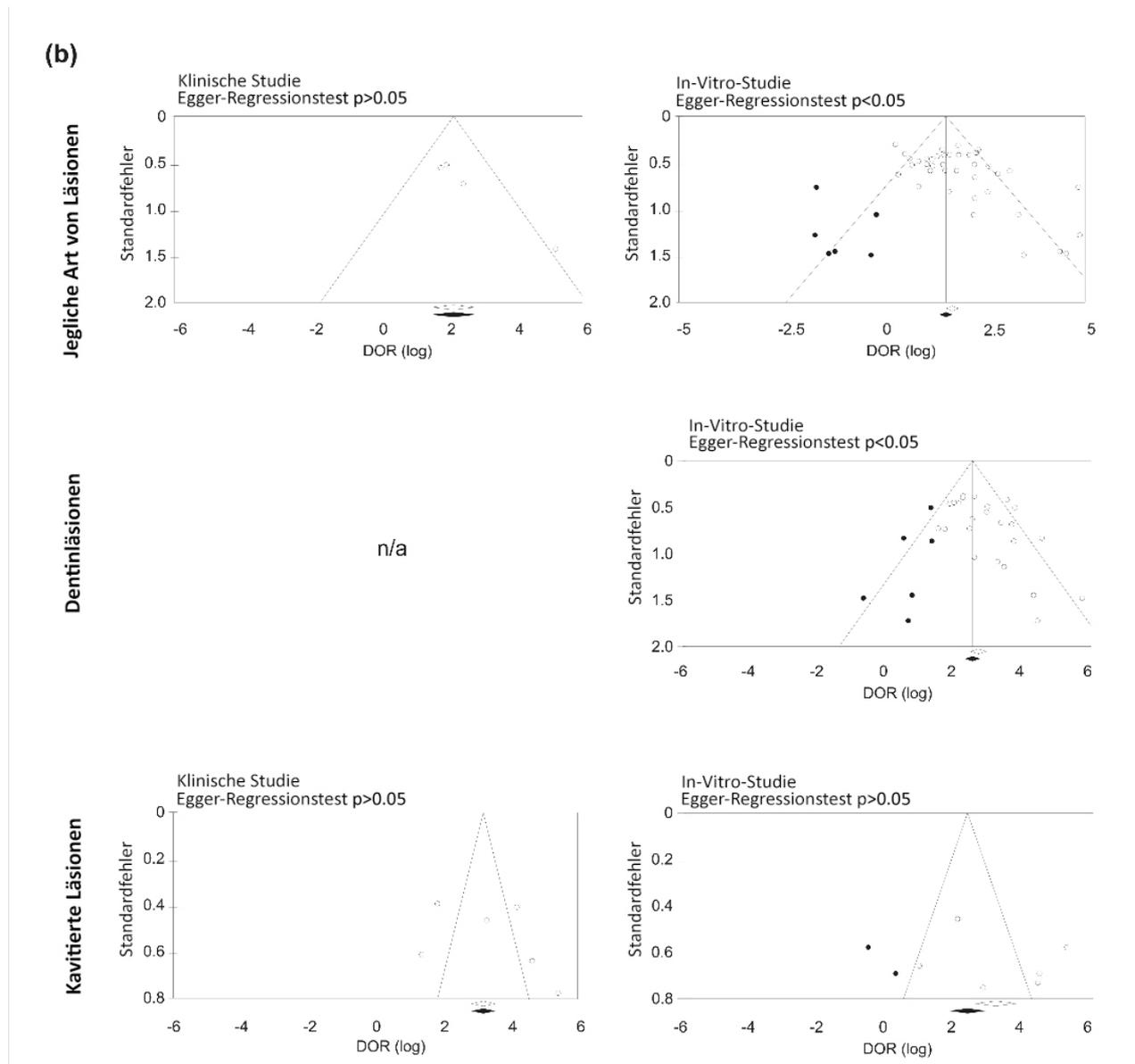


Abbildung 3-3 b: Funnel-Plots für approximale Läsionen. Die berechneten diagnostischen Quotenverhältnisse (DOR) wurden transformiert ($\log DOR$) und auf der x-Achse gegen ihren Standardfehler (auf der y-Achse) aufgetragen. Ein asymmetrischer Trichter weist auf einen möglichen Publikationsbias oder Effekte kleiner Studien hin. Die Genauigkeit der radiologischen Kariesdetektion jeglicher Läsionen, von Dentinfläsionen und kavitierten Läsionen wurde separat untersucht. Die Läsionen wurden weiterhin in okklusale (a) und approximale Läsionen (b) unterteilt. Für drei Analysen standen weniger als drei Studien zur Verfügung und die Erstellung von Funnel-Plots war somit nicht reliierbar (n/a). Weißer Diamant: gepoolte DOR basierend auf veröffentlichten Studien. Schwarzer Diamant: gepoolte DOR einschließlich geschätzter Werte. (Schwendicke et al., 2015).

3.5 Variabilität und Metaregression

Zur weiteren Beurteilung der Variabilität der Genauigkeit der röntgenologischen Kariesdetektion wurden der jeweils beste und schlechteste Datensatz in Bezug auf Sensitivität und Spezifität mit den zugehörigen Werten zur DOR und zur Heterogenität miteinander verglichen. Für alle Vergleiche waren sowohl Sensitivität als auch Spezifität höher bei der besten Genauigkeit verglichen mit der schlechtesten Genauigkeit. Insgesamt lag die Genauigkeit (DOR) bei den besten diagnostischen Umständen/Kovariaten um 23 bis 123 % höher verglichen mit den schlechtesten diagnostischen Umständen/Kovariaten (zum Beispiel bester versus schlechtester Untersucher, bestes versus schlechtestes Detektions-/Bewertungssystem), was darauf schließen lässt, dass hier eine große Variabilität innerhalb der einzelnen Studien besteht. Die Werte der Heterogenität waren in diesen separaten Analysen ähnlich jenen in der Hauptanalyse.

Uni- und multivariate Metaregressionsanalysen wurden durchgeführt, um den Einfluss potenzieller Störfaktoren (Jahr der Veröffentlichung, Verzerrungspotenzial, Anwendbarkeit [niedriges oder hohes/unklares Risiko], Dentition [Milchzähne versus bleibende Zähne] zu ermitteln. Ausschließlich das Jahr der Veröffentlichung zeigte einen signifikanten Einfluss auf die DOR für den Cut-off der Detektion kavierter Läsionen an Approximalflächen unter In-vitro-Bedingungen (R^2 [95 % CI]: -0,9 [-0,13/-0,05]). Wie bereits erwähnt, verwendeten alle Studien bis auf eine Ausnahme als Schwellenwert für das Vorliegen einer Kavität die röntgenologische Ausdehnung der Läsion bis in das Dentin. Da sich die Genauigkeit für diesen Schwellenwert über die Zeit aufgrund einer sich verändernden Läsionsaktivität ebenfalls verändert haben könnte, wurden alle betreffenden Studien ebenfalls einer Metaregressionsanalyse unterzogen. Das Jahr der Veröffentlichung wurde hierbei als Einflussgröße verwendet. In diesem Fall jedoch zeigte sich kein signifikanter Einfluss auf die Metaregression (Abb. 3-4).

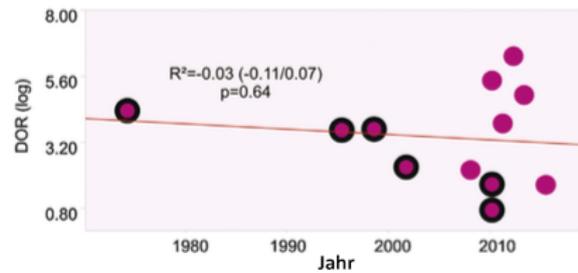


Abbildung 3-4: Metaregressionsanalyse. Bei der Analyse der Genauigkeit der Detektion kavierter approximaler Läsionen wurde das Quotenverhältnis (DOR) mit dem Jahr der Veröffentlichung verknüpft, um eine mögliche Änderung der Genauigkeit des zur Anwendung gekommenen Schwellenwertes der kavitierten Läsion (Läsion bis in das Dentin) zu bewerten. Die Meta-Regression konnte eine derartige Änderung nicht bestätigen ($p < 0,05$). Die Steigung der Regressionslinie (R^2) und die 95%-Konfidenzintervalle werden hier dargestellt. In-vitro-Studien sind zusätzlich durch schwarze Ringe gekennzeichnet. (Schwendicke et al., 2015).

4 Diskussion

Neben der visuell-taktilen Diagnostik stellt das Bissflügelröntgen zur Kariesdetektion in der täglichen zahnärztlichen Praxis das am häufigsten angewendete Diagnostikverfahren dar (Rindal et al., 2010). In Anbetracht der Tatsache, dass die Anwendung dieses Hilfsmittels für den Patienten jedoch mit der Aufnahme potenziell schädlicher ionisierender Strahlung einhergeht und gleichzeitig eine zunehmende Anzahl alternativer Verfahren zur Verfügung steht, welche geeignet sind, sowohl initiale als auch fortgeschrittene okklusale Läsionen zu detektieren, sollte die Genauigkeit des Bissflügelröntgens zur Kariesdiagnostik kritisch hinterfragt werden. Weiterhin gilt es zu hinterfragen, ob jene Genauigkeit in Abhängigkeit von Stadium und Lokalisierung der zu detektierenden Läsionen schwankt, was einen Einfluss auf die Indikationsstellung dieses Diagnostikverfahrens hätte. Aus methodischer Sicht ist zudem relevant, ob Studien mit verschiedenen diagnostischen Umgebungen (in vitro oder klinisch) und Testmethoden (z. B. Goldstandards) die röntgenologische Diagnostik unterschiedlich bewerten, da dies für die Gestaltung eines zukünftigen „optimalen“ Studiendesigns von Bedeutung wäre (und zudem Rückschlüsse auf die Validität des vorhandenen Studienkörpers hätte). Unter Berücksichtigung der großen Anzahl veröffentlichter Studien und der großen Variabilität der jeweils gewonnenen Erkenntnisse liegt es nahe, eine systematische Übersichtsarbeit und Metaanalyse anzufertigen, um die zur Verfügung stehenden Daten zu sammeln und auszuwerten und somit einige der vorgenannten Fragestellungen beantworten zu können.

Die hier vorliegende Metaanalyse von Studien, die Angaben zur Genauigkeit des Bissflügelröntgens zur Kariesdiagnostik beinhalten, machte deutlich, dass dieses Verfahren eine beschränkte Sensitivität bei der Detektion jeglicher Läsionen (initiale Läsionen und Dentinläsionen) an Okklusalfächen und, mehr noch, im approximalen Bereich aufweist. Im Gegensatz dazu war die Spezifität des Bissflügelröntgens generell sowohl im approximalen Bereich als auch okklusal hoch. Dieses Diagnostikverfahren scheint zudem insbesondere für die Detektion vorangeschrittener und vor allem kavierter Läsionen im approximalen Bereich geeignet zu sein. Sowohl Sensitivität als auch Spezifität waren hier hoch und die resultierende Genauigkeit entsprechend ebenso. Diese Erkenntnisse befinden sich im Einklang mit Ergebnissen vergangener Veröffentlichungen (Bader et al., 2001, Wenzel et al., 2004). Der Vergleich von Studienergebnissen, welche unter In-vitro-Bedingungen beziehungsweise klinischen Bedingungen gewonnen wurden, zeigte, dass erstere möglicherweise zu einer Überschätzung der Sensitivität und einer Unterschätzung der Spezifität führen. Zudem scheinen jene Studien noch anfälliger für einen Publikationsbias und das assoziierte Verzerrungspotenzial zu sein.

Die Genauigkeit des Bissflügelröntgens sollte auch im Vergleich mit alternativen Kariesdiagnostikverfahren, wie zum Beispiel auf Fluoreszenz basierenden Verfahren, diskutiert werden. In Bezug auf Okklusalfächen wiesen letztere Verfahren bei der Detektion initialer Läsionen eine signifikant höhere Genauigkeit, Sensitivität und Spezifität auf (Gimenez et al., 2013). Allerdings wich die Definition „initialer Läsionen“ zwischen dieser und der vorliegenden Arbeit voneinander ab. Denn in der hier vorliegenden Arbeit wurden die drei Schwellenwerte „jegliche Läsionen“, „Dentinläsionen“ und „kavitierte Läsionen“ gewählt, wobei initiale Läsionen somit unter den Schwellenwert „jegliche Läsionen“ fallen und in diesem Sinne keine eigene Kategorie darstellen. Bei weiter fortgeschrittenen okklusalen Läsionen schien das Bissflügelröntgen weniger sensitiv, jedoch aber spezifischer als auf Fluoreszenz basierenden Maßnahmen zu sein, sowohl bezogen auf bleibende Zähne als auch auf Milchzähne. Bezogen auf die Diagnostik an Approximalfächen bleibender Zähne stehen für röntgenologische Verfahren deutlich mehr Studien zur Verfügung, als für auf Fluoreszenz basierenden Verfahren (lediglich zwei Studien wurden diesbezüglich im letzten aktuellen Review inkludiert). Für Milchzähne ist die Zahl der Studien zu beiden Verfahren hingegen ähnlich; möglicherweise wurden hier auf Fluoreszenz basierende Verfahren häufiger getestet, da eine röntgenologische Kariesdetektion im Milchgebiss und damit bei kleineren Kindern aus o. g. Gründen oft nur restriktiv angezeigt ist und strahlungsfreie Methoden als mögliche Alternativen zum Röntgen von großem Interesse sind. Auch hier schienen auf Fluoreszenz basierende Maßnahmen sensitiver, aber signifikant weniger spezifisch, einhergehend mit einer niedrigeren Genauigkeit. Für den Schwellenwert der kavitierten Läsionen, welcher letztlich maßgeblichen Einfluss auf die Therapieentscheidung hat, waren für auf Fluoreszenz basierende Maßnahmen keinerlei systematisch analysierte Daten verfügbar. Hingegen weisen die in der vorliegenden Arbeit gesammelten Daten auf eine hohe Genauigkeit des Bissflügelröntgens zur Detektion kavittierter Läsionen hin.

Im Vergleich zur visuell-taktilen Kariesdiagnostik scheint das Bissflügelverfahren bei der Diagnostik von Schmelz- und Dentinläsionen oder kleinen kavitierten approximalen Läsionen überlegen zu sein, da diese weder dem Auge noch der Sonde des Behandlers zugänglich sind, während die visuell-taktile Kariesdiagnostik aus offensichtlich nachvollziehbaren Gründen gut geeignet ist, kavitierte okklusale, Glattflächenläsionen und sehr tiefe proximale Läsionen zu diagnostizieren (Bader et al., 2001).

Zusammengefasst lässt sich sagen, dass das Bissflügelröntgen insbesondere geeignet ist, um Dentinläsionen und vor allem kavitierte proximale Läsionen zu detektieren, wobei hier eine geringere Sensitivität, jedoch aber eine hohe Spezifität bei der Diagnostik initialer kariöser Läsionen vorliegt. Die Detektion letzterer Läsionen sollte non- oder mikro-invasive Behandlungsmaßnahmen nach sich ziehen (wie zum Beispiel Remineralisation, Fissurenversiegelung oder Infiltration), noch

bevor es unbehandelt zu einem Oberflächeneinbruch kommen kann (Petersson et al., 2004). Während auf Fluoreszenz basierende Maßnahmen besser geeignet sind, solche frühen Läsionen zu diagnostizieren, so weisen sie in diesem Fall jedoch eine geringere Spezifität auf, was mit der Gefahr einer Überdetektion einhergeht (falsch positive Detektionen). Dies wiederum wäre unproblematisch, wenn die resultierende Überbehandlung grundsätzlich in einer non- oder mikro-invasiven und nicht einer invasiven Therapie bestehen würde (Schwendicke et al., 2014). Da jedoch die Mehrheit der Zahnärzte dazu neigt, bereits bei röntgenologisch auf den Zahnschmelz begrenzten Läsionen eine restaurative Therapie durchzuführen (Gomez et al., 2014; Gordan et al., 2009; Kakudate et al., 2012; Traebert et al., 2007; Vidnes-Kopperud et al., 2011), sollte die Spezifität eine höhere Priorität als der Wert der Sensitivität bei der Detektion initialer Läsionen genießen. Dies ist insbesondere daher wichtig, da wir derzeit einen allgemeinen Rückgang der Kariesprävalenz beobachten können, wobei hierbei einige wenige (häufig sozioökonomisch) benachteiligte Schichten die Mehrheit der kariösen Läsionen in sich tragen und andererseits viele (vor allem junge) Patienten nahezu gar keine kariösen Läsionen aufweisen (Marthaler et al., 1996; Marthaler TM, 2004; Schwendicke et al., 2015). Innerhalb dieser Patientengruppe ist der positive Vorhersagewert für auf Fluoreszenz basierende Detektionsverfahren mit einer geringen Spezifität bei der Detektion initialer Läsionen ebenfalls gering.

Im Gegensatz zur Detektion initialer Läsionen scheint das Bissflügelverfahren für die Detektion von Dentinläsionen an Okklusalfächen (auch als „versteckte Karies“ bekannt) gut geeignet zu sein (Ricketts et al., 1997). Da eine Ausdehnung bis in das Dentin ein Hinweis für das Vorliegen einer Kavität sein kann und somit invasive Therapieformen erforderlich macht, kommt diesem Schwellenwert eine besondere Bedeutung zu. In Bezug hierauf konnten wertvolle Informationen aus Studien entnommen werden, welche sich mit dem diagnostischen Wert von Röntgenaufnahmen befassten, bei denen es zu entscheiden galt, ob bei approximalen kariösen Läsionen ein Oberflächeneinbruch vorliegt oder nicht. Da bis auf eine einzige Ausnahme alle Studien es von diesem Schwellenwert abhängig machten, ob eine Läsion als kavitiert einzuschätzen ist oder nicht, wurde eine Metaregressionsanalyse durchgeführt, um herauszufinden, ob sich der diagnostische Wert dieses Schwellenwertes über die Zeit ändert. Dies ist insofern relevant, als dass sich die Läsionsaktivität durch den mittlerweile weitverbreiteten Einsatz von Fluoriden verringert haben könnte. Denn der lokale Fluorideinsatz führt an einer kariösen Läsion zur Aufnahme von Mineralien und hat somit zur Folge, dass die sogenannte pseudointakte Oberflächenschicht dicker wird. Somit wird das Voranschreiten des kariösen Zerfalls arretiert. Diese Arretierung kann somit auch dazu führen, dass tiefgelegene Anteile von kariösem Dentin remineralisieren (Meyer-Lückel et al., 2012).

Dies hätte zur Folge, dass kariöse Läsionen heutzutage später einen Oberflächeneinbruch erleiden, als es früher der Fall war. Jedoch ergab sich bei eben jener Analyse kein signifikanter Trend. Hieraus wiederum lässt sich schlussfolgern, dass möglicherweise diagnostizierte Oberflächeneinbrüche bei radiologischen Dentinläsionen kritisch zu bewerten sind. Somit lässt sich zusammengefasst sagen, dass vor allem in Populationen mit einem geringen Kariesrisiko noch vor der Einleitung einer invasiven Therapie über eine temporäre Zahnseparation und natürlich eine suffiziente visuell-taktile Diagnostik nachgedacht werden sollte (Wenzel et al., 2004; Ratledge et al., 2001; Wenzel et al., 1998).

Es existieren noch weitere Störfaktoren, welche einen Einfluss auf die Genauigkeit des Bissflügelröntgens zur Kariesdiagnostik haben. Einige jener Faktoren wurden in der hier vorliegenden Arbeit mittels Subgruppen- und Metaregressionsanalysen ausgewertet. In der Mehrheit der ausgewerteten Studien wurden Detektionsmethoden an bleibenden Zähnen getestet und obwohl sich diesbezüglich kein signifikanter Unterschied der Genauigkeitswerte zwischen den beiden Dentitionen nachweisen ließ, so sollte dies nicht etwa mit einem nicht vorhandenen Einfluss verwechselt, sondern vielmehr als potenzieller Mangel an statistischer Nachweiskraft verstanden werden. Weiterhin wurde der Einfluss des jeweils verwendeten Detektionssystems untersucht, also digitale Speicherfolien/Sensoren oder analoge Filme. Bei den digitalen Verfahren kommen entweder Halbleitersensoren auf CCD-Basis bzw. CMOS-Basis oder sogenannte Speicherfolien bzw. Lumineszenzfolien zur Anwendung. Als Surrogatparameter für die Abgrenzung neuerer gegen ältere Verfahren wurde das Veröffentlichungsjahr der jeweiligen Studie verwendet. Dies geschah unter der Annahme, dass in neueren Studien vermutlich modernere Röntgensysteme Verwendung gefunden haben könnten. Jedoch scheint der Einfluss des Veröffentlichungsjahres in gewisser Weise für sich genommen eine Art Störfaktor darzustellen, da jüngere Studien zum Teil ein strukturierteres Studiendesign mit einem geringeren Verzerrungspotenzial aufweisen. Daher wurden für diese Auswertung sowohl uni- als auch multivariate Metaregressionsanalysen durchgeführt, wobei das Verzerrungspotenzial und das Veröffentlichungsjahr hierbei als unabhängige Variablen verwendet wurden. Da hierbei kein signifikanter Effekt zu verzeichnen war, scheint die Anwendung neuerer Systeme im Vergleich zu älteren Systemen nicht zu einer höheren Genauigkeit zu führen, was die Erkenntnisse früherer Untersuchungen zu bestätigen scheint (Wenzel et al., 2004). Jedoch könnten die bei dieser Metaregressionsanalyse erlangten Ergebnisse durch ein bei jüngeren Studien besser strukturiertes Studiendesign oder eine realitätsnähere Kariesprävalenz begründet sein. Bei älteren Studien lag zum Teil eine Kariesprävalenz von 90 % und mehr vor, was sehr unrealistisch ist und somit eine signifikante Verzerrung darstellt. Denn eine hohe Kariesprävalenz korreliert mit einer

Erhöhung des positiven prädiktiven Wertes, was wiederum die Wahrscheinlichkeit auf ein positives Testergebnis erhöht.

Unterschiede beim Studiendesign, dem Indextest sowie der Erfahrung und der Anzahl der Untersucher wurden bereits in früheren Arbeiten als Ursachen für die Heterogenität der Studienergebnisse benannt, was durch die hier vorliegende Arbeit bestätigt wird (Bader et al., 2001). Weiterhin könnte das Vorliegen von Heterogenität auch einen Hinweis für einen potenziellen Mangel an Reliabilität (also denjenigen Anteil an Variabilität, verursacht durch tatsächliche Unterschiede im zu messenden Merkmal) darstellen. Dies wurde überprüft, indem innerhalb jeder Studie der Datensatz mit den jeweils besten und schlechtesten Werten der Genauigkeit aufgenommen wurde. Hierbei ergab sich, dass die Genauigkeit (DOR) um bis zu 123 % variierte, je nachdem, welcher Datensatz eingesetzt wurde. Es sollte hervorgehoben werden, dass dies auf einen unterschiedlichen Erfahrungsstand einzelner Untersucher oder aber auf die jeweils unterschiedlichen zur Anwendung kommenden Detektionssysteme (digital versus analog) zurückzuführen sein könnte (Bader et al., 2001). Da es sich bei der Reliabilität um einen wichtigen Parameter bei der Bewertung eines Detektionssystems handelt, sollten zukünftige Studien auf ein möglichst standardisiertes Studiendesign zurückgreifen. Dies würde die Variabilität zwischen den einzelnen Veröffentlichungen verringern und eine Abschätzung der Reliabilität jenseits von rein studienmethodischen Schwankungen ermöglichen. Weiterhin sollte die von den einzelnen Untersuchern mit jeweils unterschiedlichem Erfahrungsstand erzielte Genauigkeit, also jeweils Erfahrungsstand und zugehörige Genauigkeitsangabe, detailreicher angegeben werden. Schließlich sollten sich neue Standards für Studiendesigns auch auf die Validationskriterien beziehen, da sich auch dies auf die jeweils gemessene Genauigkeit auswirkt. Die Metaregressionsanalyse hat dabei keinen signifikanten Einfluss der unterschiedlichen Validationskriterien (also destruktive versus nicht destruktive) auf die gemessene Genauigkeit zutage gefördert, jedoch könnte es sich allerdings auch in diesem Fall potenziell eher um einen Mangel an statistischer Nachweiskraft als einen tatsächlichen Beweis handeln. Denn bei der Anwendung nicht destruktiver Validationskriterien werden zwangsläufig falsch-negative Ergebnisse ausgeblendet, da hier im Gegensatz zur Anwendung destruktiver Validationskriterien nicht sämtliche Zahnoberflächen histologisch beziehungsweise mikroradiografisch untersucht werden können. Im Falle klinischer Studien ist es bis auf ganz bestimmte Ausnahmen (wie zum Beispiel die temporäre Separation mit kieferorthopädischen Spanngummis) aus ethischen Gründen nicht möglich, jede Zahnoberfläche invasiv zu validieren, bei In-vitro-Studien jedoch schon.

Das Hauptaugenmerk bei der Einschätzung der Güte des Bissflügelröntgens zur Detektion von Karies auf dem Kriterium der Validität, das heißt Genauigkeit verglichen gegen einen Goldstandard, wurde bereits früher kritisiert (Baelum et al., 2006). Denn zum einen besteht kein Konsens bezüglich eines ganz konkreten Goldstandards und zum anderen könnte dies für den Patienten ohnehin irrelevant sein. Für ihn ist letztlich die prognostische Validität (und in noch viel größerem Maße der gesundheitliche Nutzen) eines Tests von größerer Bedeutung als seine Genauigkeit (Schwendicke et al., 2014; Baelum et al., 2006). In diesem Sinne muss das Augenmerk auf den Parameter der Genauigkeit in der hier vorliegenden Arbeit in gewisser Weise als Einschränkung angesehen werden. Eine weitere Einschränkung besteht in der Qualität der Berichte über den Validationsprozess in den ausgewählten Veröffentlichungen. Die überwiegend leider niedrige Qualität der inkludierten Studien könnte einen Einfluss auf die Gültigkeit der abgeleiteten Schlussfolgerungen haben und sollte demnach hervorgehoben werden. So könnte beispielsweise die Untersuchung nicht repräsentativer Zahnpopulationen, wie zum Beispiel Weisheitszähne oder aus kieferorthopädischen Gründen extrahierte Prämolaren, zu Verzerrungen führen. Auch ein nicht konsistenter zeitlicher Testverlauf (mangelnde Verblindung durch zu kurze Abstandszeiten zwischen Index- und Referenztest) könnte einen Einfluss auf Ergebnisse der jeweils ermittelten Genauigkeiten haben. Auch könnte die sogenannte Cluster-Bildung (oder Gruppierung) eine weitere Einflussgröße darstellen. Dieses Phänomen bezieht sich auf die Patientenauswahl in den einzelnen hier ausgewerteten Studien. Die Patientenauswahl sollte immer zufällig erfolgen. Dabei sollten weder sämtliche untersuchte Patienten aus einer einzelnen Zahnarztpraxis stammen, noch sollten mehrere Zähne eines einzelnen Patienten untersucht worden sein. Es handelt sich hierbei andernfalls um eine statistische Häufung, welche gleichermaßen in eine Gruppenbildung der korrelierenden Untersuchungsergebnisse mündet. Dieses, als Cluster-Bildung bezeichnete Phänomen, führt letztlich zur Verringerung der statistischen Nachweiskraft und zu fälschlicherweise engen Konfidenzintervallen (je breiter das Konfidenzintervall ist, desto unpräziser ist eine Schätzung), also in gewisser Weise zu einer vorgetäuschten Genauigkeit (Litaker et al., 2013). Ein weiteres Problem besteht in der Bandbreite der in den einzelnen Veröffentlichungen jeweils angewendeten unterschiedlichen Detektions- und Evaluationssystemen (analog, digital, neuere oder ältere analoge beziehungsweise digitale Systeme etc.). Denn aufgrund von unvollständigen Angaben und auch aufgrund der Vielzahl der unterschiedlichen Detektionshilfsmittel in den einzelnen Studien war es oftmals nicht möglich, die unterschiedlichen Methoden innerhalb von Subgruppenanalysen zu vergleichen. Auch schwankte die Anzahl der zur Subgruppenanalyse herangezogenen Studien und die Mehrzahl der Studien wurde unter In-Vitro-Bedingungen und nicht unter klinischen Bedingungen durchgeführt. Letzteres ist verständlich, da eine suffiziente (invasive) Validation aus ethischen Gründen unter klinischen

Bedingungen nicht realisierbar ist, wie oben dargelegt. Allerdings sollte hervorgehoben werden, dass weder in der hier vorliegenden Arbeit noch einem anderen Review ein maßgeblicher Unterschied zwischen den Genauigkeitsangaben von In-vitro-Studien beziehungsweise klinischen Studien festgestellt werden konnte (Bader et al., 2001). Schließlich muss auch noch auf Limitationen der hier vorliegenden Arbeit selbst hingewiesen werden. So könnte zum Beispiel der hier angewendete Suchalgorithmus zur Identifikation potenziell relevanter Studien nicht umfassend genug gewesen sein. Erschwerend kommt hinzu, dass Studien, welche sich mit dem Thema der Genauigkeit von Röntgenverfahren befassen, oftmals nicht klar in den entsprechenden Datenbanken für die Literatursuche als solche kenntlich gemacht sind (Akers et al., 2009). Weiterhin ist die Datenextraktion für jene Art von Studien anfällig für Verzerrungen, da die Berichtsqualität, wie dargelegt, niedrig ist und teilweise auch unklar ist, ob alle erzielten Ergebnisse berichtet wurden (selektives Berichten). Schlussendlich sei erwähnt, dass das Augenmerk in der hier vorliegenden Arbeit auf Primärkaries gelegt wurde. Unter dem Aspekt, dass Sekundärkaries (oder auch Füllungsrandkaries) eine der häufigsten Indikationen für invasive Maßnahmen darstellt, sollte in Zukunft weiter in dieser Richtung geforscht werden (Fontana et al., 2000).

Abschließend kann bemerkt werden, dass das Bissflügelröntgen zur Kariesdetektion nach dem aktuellen Forschungsstand relativ wenig sensitiv, wohl aber hoch spezifisch für die Detektion initialer okklusaler und approximaler Läsionen ist. Sensitivität, Spezifität und Genauigkeit sind im Vergleich zu jenen frühen Läsionen signifikant höher für kariöse Defekte, welche sich bis in das Dentin erstrecken, beziehungsweise für kavitierte proximale Läsionen. Die Mehrheit der inkludierten Studien wies ein hohes Verzerrungspotenzial auf; zudem sind die Ergebnisse beispielsweise aufgrund von In-Vitro-Bedingungen nicht immer auf die allgemeine zahnärztliche Praxis zu übertragen. Dies scheint jedoch im Rahmen der hier angewendeten statistischen Auswertung keinen Einfluss auf die Werte der Genauigkeitsangaben gehabt zu haben.

Beim Design zukünftiger Studien sollte nicht ausschließlich darauf Wert gelegt werden, die bereits vorhandenen Daten lediglich zu erweitern beziehungsweise zu ergänzen. Vielmehr sollte über die Etablierung eines möglichst standardisierten Studienaufbaus nachgedacht werden, wobei nach Möglichkeit auch auf repräsentativere Studienpopulationen zurückgegriffen werden sollte. Auch sollte bei künftigen Arbeiten auf eine suffiziente Verblindung der Untersucher geachtet werden, wobei zwischen der Durchführung von Index- und Referenztest ein entsprechendes zeitliches Intervall bedacht werden muss. Die Einbeziehung von Untersuchern mit einem jeweils unterschiedlichen Erfahrungsstand ist ebenfalls wünschenswert. Jedoch sollte hierbei beachtet werden, dass in diesem Falle auch konkret die einzelnen Ergebnisse für jeden Untersucher zugeordnet

sein müssen. Die in den einzelnen Studien verwendeten Schwellenwerte müssen im Detail erklärt werden. Weiterhin sollte angestrebt werden, sowohl auf allgemeingültige Schwellenwerte als auch Referenztests zurückzugreifen. Schließlich sollte sowohl in der täglichen zahnärztlichen Praxis als auch im Bereich der wissenschaftlichen Forschung das Hauptaugenmerk nicht ausschließlich auf die Genauigkeit eines Diagnoseverfahrens gelegt werden, sondern auch auf den resultierenden gesundheitlichen Vorteil eines jeden Individuums innerhalb verschiedener Populationen.

Literaturverzeichnis

1. Bader J, Shugars D, Bonito A (2001): Systematic reviews of selected dental caries diagnostic and management methods. *Journal of Dental Education*, 65:960–8.
2. Gordan VV, Garvan CW, Richman JS, Fellows JL, Rindal DB, Qvist V, Heft MW, Dale Williams O, Gilbert GH (2009): How dentists diagnose and treat defective restorations: evidence from the dental practice-based research network. *Operative Dentistry*,34:664–73.
3. Rindal DB, Gordan VV, Litaker MS, Bader JD, Fellows JL, Qvist V, Wallace-Dawson MC, Anderson ML, Gilbert GH (2010): Methods dentists use to diagnose primary caries lesions prior to restorative treatment: findings from The Dental PBRN. *Journal of Dentistry*, 38:1027–32.
4. Statistisches Jahrbuch der Bundeszahnärztekammer (2013/14).
5. Fejerskov O, Nyvad B, Kidd E (2003): Clinical and radiographic caries diagnosis; in Fejerskov O, Kidd EA (eds.): *Dental Caries – The disease and its management*. ec. Oxford: Blackwell Munksgaard. pp 111–128.
6. Gimenez T, Braga MM, Raggio DP, Deery C, Ricketts DN, Mendes FM (2013): Fluorescence-based methods for detecting caries lesions: systematic review, meta-analysis and sources of heterogeneity. *PLOS ONE*, 8:e60421.
7. WPGS (2019)
8. Schmidt R & Factor R (2013): Understanding Sources of Bias in Diagnostic Accuracy Studies, *Archives of Pathology & Laboratory Medicine*, Vol. 137.

9. Schwendicke F, Stolpe M, Meyer-Lueckel H, Paris S (2014): Detecting and treating occlusal caries lesions: a cost-effectiveness analysis, *Journal Dental Research*, <http://dx.doi.org/10.1177/0022034514561260>
10. Dritte deutsche Mundgesundheitsstudie (1997)
11. Vierte deutsche Mundgesundheitsstudie (2006)
12. Marthaler T, O'Mullane D, Vrbic V (1996): The prevalence of dental caries in Europe 1990–95, *Caries Research*, 30:237–55.
13. Marthaler TM (2004): Changes in dental caries 1953–2003, *Caries Research*, 38:173–81.
14. Moher D, Liberati A, Tetzlaff J, Altman D (2009): Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement, *PLoS Medicine*, 6:e1000097.
15. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy.
16. Preferred reporting items for systematic reviews and meta-analyses.
17. Meyer-Lückel H, Paris S, Ekstrand K (2012): *Karies – Wissenschaft und klinische Praxis*.
18. Deville W, Bezemer P, Bouter L (2000): Publications on diagnostic test evaluation in family medicine journals: an optimal search strategy, *Journal of Clinical Epidemiology*, 53:65–9.
19. Higgins JPT, Green S, (eds.) (2011): *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.10 (updated March 2011), The Cochrane Collaboration.

20. Whiting P, Rutjes A, Westwood M, Mallett S, Deeks J, Reitsma J, Leeflang MGM, Sterne JAC, Bossuyt PMM, QUADAS-2 Group (2011): QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies, *Annals of Internal Medicine*, 155:529–36.
21. Glas AS, Lijmer JG, Prins MH, Bossel GJ, Bossuyt PMM (2003): The diagnostic odds ratio: a single indicator of test performance, *Journal of Clinical Epidemiology*, 56:1129–35.
22. Hanley JA, McNeil BJ (1982): The meaning and use of the area under a receiver operating characteristic (ROC) curve, *Radiology*, 143:29–36.
23. Walter SD (2002): Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data, *Statistics in Medicine*, 21:1237–56.
24. Higgins JPT, Thompson SG (2002): Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21:1539–58.
25. Egger M, Smith GD, Schneider M, Minder C (1997): Bias in meta-analysis detected by a simple, graphical test, *British Medical Journal*, 315:629–34.
26. Duval S & Tweedie R (2000): Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56:455–63.
27. Sansare K, Singh D, Sontakke S, Karjodkar F, Saxena V, Frydenberg M, Wenzel A (2014): Should cavitation in proximal surfaces be reported in cone beam computed tomography examination? *Caries Research*, 48:208–13.
28. Wenzel A (2004): Bitewing and digital bitewing radiography for detection of caries lesions, *Journal Dental Research*, 83:C72–5.
29. Petersson LG, Twetman S, Dahlgren H, Norlund A, Holm A-K, Nordenram G, Lagerlöf F, Söder B, Källestal C, Mejare I, Axelsson S, Lingström P (2004):

- Professional fluoride varnish treatment for caries control: a systematic review of clinical trials, *Acta Odontologica Scandinavica*, 62:170–176.
30. Gomez J, Ellwood RP, Martignon S, Pretty IA (2014): Dentists' perspectives on caries-related treatment decisions, *Community Dental Health*, 31:91–8.
31. Gordan VV, Garvan CW, Heft MW, Fellows JL, Qvist V, Rindal DB, Gilbert GH (2009): Restorative treatment thresholds for interproximal primary caries based on radiographic images: findings from the Dental Practice-Based Research Network, *General Dentistry*, 57:654–63; quiz 64–6, 595, 680.
32. Kakudate N, Sumida F, Matsumoto Y, Manabe K, Yokoyama Y, Gilbert GH, Gordan VV (2012): Restorative treatment thresholds for proximal caries in dental PBRN, *Journal Dental Research*, 91:1202–8.
33. Traebert J, Wesolowski CI, de Lacerda JT, Marcenes W (2007): Thresholds of restorative decision in dental caries treatment among dentists from small Brazilian cities, *Oral Health & Preventive Dentistry*, 5:131–5.
34. Vidnes-Kopperud S, Tveit AB, Espelid I (2011): Changes in the treatment concept for approximal caries from 1983 to 2009 in Norway, *Caries Research*, 45:113–20.
35. Schwendicke F, Dörfer C, Schlattmann P, Foster Page L, Thomson M, Paris S (2015): Socioeconomic inequality and caries: a systematic review and meta-analysis, *Journal Dental Research*, 94:10–8.
36. Ricketts D, Kidd E, Weerheijm K, de Soet H (1997): Hidden caries: what is it? Does it exist? Does it matter? *International Dental Journal*, 47:259–65.
37. Ratledge DK, Kidd EA, Beighton D (2001): A clinical and microbiological study of approximal carious lesions, Part 1: the relationship between cavitation, radiographic

- lesion depth, the site-specific gingival index and the level of infection of the dentine. *Caries Research*, 35:3–7.
38. Wenzel A, Hintze H, Horsted-Bindslev P (1998): Accuracy of radiographic detection of residual caries in connection with tunnel restorations, *Caries Research*, 32:17–22.
39. Baelum V, Heidmann J, Nyvad B (2006): Dental caries paradigms in diagnosis and diagnostic research, *European Journal of Oral Sciences*, 114:263–77.
40. Litaker MS, Gordan VV, Rindal DB, Fellows JL, Gilbert GH (2013): Cluster Effects in a National Dental PBRN restorative study, *Journal of Dental Research*, 92(9): 782–787.
41. Akers J (2009): Centre for Reviews Dissemination. *Systematic Reviews: CRD's guidance for undertaking reviews in health care*. York: York Publishing Services.
42. Fontana M & González-Cabezas C (2000): Secondary caries and restoration replacement: an unresolved problem, *Compendium of continuing education in dentistry*, 21(1):15–8, 21–4, 26 passim; quiz 30.

Eidesstattliche Versicherung

„Ich, Markus Tzschoppe, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: „Sensitivität und Spezifität intraoraler Röntgenaufnahmen zur Detektion von Karies – eine systematische Metaanalyse und Übersichtsarbeit / Radiographic caries detection: A Systematic Review and Meta-Analysis“ selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.

Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren/innen beruhen, sind als solche in korrekter Zitierung kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) werden von mir verantwortet.

Meine Anteile an etwaigen Publikationen zu dieser Dissertation entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit dem/der Erstbetreuer/in, angegeben sind. Für sämtliche im Rahmen der Dissertation entstandenen Publikationen wurden die Richtlinien des ICMJE (International Committee of Medical Journal Editors; www.icmje.org) zur Autorenschaft eingehalten. Ich erkläre ferner, dass ich mich zur Einhaltung der Satzung der Charité – Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis verpflichte.

Weiterhin versichere ich, dass ich diese Dissertation weder in gleicher noch in ähnlicher Form bereits an einer anderen Fakultät eingereicht habe.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§§156, 161 des Strafgesetzbuches) sind mir bekannt und bewusst.“

Datum und Unterschrift Markus Tzschoppe

Anteilsklärung an etwaigen erfolgten Publikationen

Ich, Markus Tzschoppe, hatte Anteil an der folgenden Publikation:

Schwendicke F, Tzschoppe M, Paris S (2015): Radiographic caries detection: A Systematic Review and Meta-Analysis, Journal of Dentistry, 43;924-9.

Hiermit erkläre ich, Markus Tzschoppe, dass Teile aus meiner Monographie bereits publiziert worden sind. Ich habe die Suche in digitalen Datenbanken und grauer Literatur selbständig durchgeführt und nach entsprechender Filterung von Titeln und Abstrakten die verbliebenen Artikel im Volltext selbständig ausgewertet. Aus meiner statistischen Auswertung sind die Tabellen 3-1 und 3-2 sowie die Abbildungen 3-1, 3-2, 3-3 und 3-4 entstanden. Diese darf ich als Koautor mit Genehmigung des Elsevier Verlages in meiner Dissertation verwenden.

Datum, Unterschrift und Stempel Prof. Dr. Falk Schwendicke

Datum und Unterschrift Markus Tzschoppe

Lebenslauf/Vita

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

Komplette Publikationsliste

Publikation 1:

Veröffentlichung der Ergebnisse der Meta-Analyse unter dem Titel “Radiographic caries detection: A Systematic Review and Meta-Analysis” im Journal of Dentistry. 2015; 43;924-9.

(Autoren: Schwendicke F, Tzschoppe M, Paris S.)

Danksagung

Für seine immer hilfreiche Unterstützung und durchweg konstruktive Kritik möchte ich Herrn Prof. Dr. Schwendicke aus tiefster Überzeugung danken. Nie war eine Rückfrage zu trivial und noch viel weniger eine Zeit zu unchristlich für ein Gespräch oder ein E-Mail. Tausend Danke für Ihre treue Unterstützung vom Anfang bis zum heiß ersehnten Abschluss.

Meiner hochgeschätzten Freundin und Kollegin Anna Skopp möchte ich ebenfalls so herzlich danken, wie es nur irgend möglich ist. Wir haben uns zusammen bewaffnet mit ungefähr 1'000'000 Hektolitern edler Kaffevariationen und motivatorisch hochwirksamer & zielgerichteter Hintertritte durch die stressigen Phasen unseres Studiums und unserer Dissertationen gekämpft und uns an den richtigen Stellen gegenseitig aufgebaut und motiviert. Danke.

Abschließend möchte ich meiner Familie aus tiefstem Herzen danken. Ihr habt mir dieses Studium, die Verwirklichung all meiner Träume (auch sonst eigentlich alles in meinem Leben) und nun diesen krönenden Abschluss ermöglicht. Ihr seid und wart immer für mich da. Und zum Schluss aber nicht letztens danke ich Dir, lieber Andreas. Auf unserem Bohrturm habe ich viel von Dir gelernt und vieles davon hat den Abschluss dieser Arbeit direkt und indirekt erst möglich gemacht.