

Aus der Poliklinik für Zahnerhaltung und Parodontologie  
Klinikum der Ludwig-Maximilians-Universität München



**Zur Genauigkeit der Erkennung restaurativer  
Behandlungsmaßnahmen auf apikalen Zahnfilmen**

Dissertation

zum Erwerb des Doktorgrades der Zahnmedizin  
an der Medizinischen Fakultät der  
Ludwig-Maximilians-Universität München

vorgelegt von

Annika Wülk

aus

Würzburg

Jahr

2023

---

Mit Genehmigung der Medizinischen Fakultät der  
Ludwig-Maximilians-Universität zu München

Erster Gutachter: Prof. Dr. Jan Kühnisch

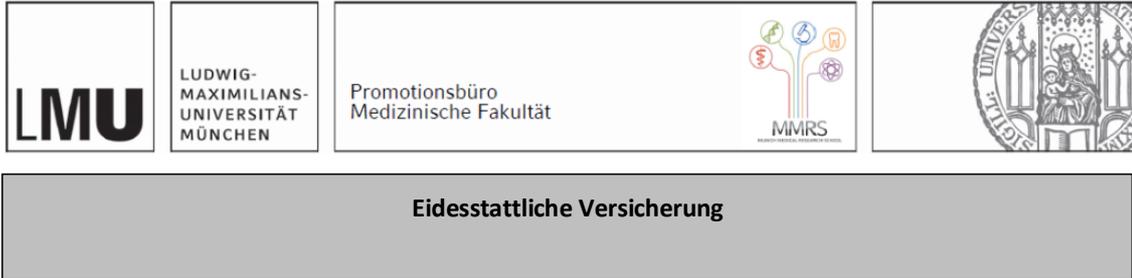
Zweiter Gutachter: Prof. Dr. Roswitha Heinrich-Weltzien

Dritter Gutachter: Prof. Dr. Bogna Stawarczyk

Dekan: Prof. Dr. med. Thomas Gudermann

Tag der mündlichen Prüfung: 21.11.2023

## Affidavit



### Eidesstattliche Versicherung

Wülk, Annika

Name, Vorname

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel:

### **„Zur Genauigkeit der Erkennung restaurativer Behandlungsmaßnahmen auf apikalen Zahnfilmen“**

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, 06.12.2023

Ort, Datum

Annika Wülk

Unterschrift Doktorandin bzw. Doktorand

## Inhaltsverzeichnis

<b>Affidavit</b> .....	<b>3</b>
<b>Inhaltsverzeichnis</b> .....	<b>4</b>
<b>Publikationsliste</b> .....	<b>5</b>
<b>1. Beitrag zu den Veröffentlichungen</b> .....	<b>6</b>
<b>2. Einleitung</b> .....	<b>7</b>
<b>3. Zielstellung</b> .....	<b>9</b>
<b>4. Material und Methoden</b> .....	<b>10</b>
<b>5. Ergebnisse</b> .....	<b>12</b>
<b>6. Diskussion</b> .....	<b>13</b>
<b>7. Zusammenfassung und Ausblick</b> .....	<b>16</b>
<b>8. Englische Zusammenfassung (Abstract)</b> .....	<b>17</b>
<b>9. Veröffentlichung I</b> .....	<b>19</b>
<b>10. Veröffentlichung II</b> .....	<b>20</b>
<b>11. Literaturverzeichnis</b> .....	<b>21</b>
<b>Danksagung</b> .....	<b>25</b>

## Publikationsliste

Wülk, A., Meusbürger, T., Dujic, H., Hickel, R., Kessler, A., Heck, K., Kühnisch, J. (2023). The Dentist's Ability to Detect Different Restorative Procedures on Periapical Radiographs-Results from a Reliability Study. International Journal of Environmental Research and Public Health, 20(3), 2619. <https://doi.org/10.3390/ijerph20032619>

Meusbürger, T., Wülk, A., Kessler, A., Heck, K., Hickel, R., Dujic, H., Kühnisch, J. (2023). The Detection of Dental Pathologies on Periapical Radiographs-Results from a Reliability Study. Journal of Clinical Medicine, 12(6), 2224. <https://doi.org/10.3390/jcm12062224>

# 1. Beitrag zu den Veröffentlichungen

Im Rahmen des Dissertationsprojektes waren die nachfolgenden Arbeitsschritte notwendig, die wie folgt bearbeitet wurden.

## 1.1 Publikation: "The dentist's ability to detect different restorative procedures on periapical radiographs - Results from a reliability study"

	Annika Wülk	Prof. Dr. Kühnisch	Co-Autoren
Projektidee/Studiendesign	25%	50%	25%
Literaturrecherche	100%	-	-
Bildselektion	100%	-	-
Training und Kalibrierung der Arbeitsgruppe	25%	50%	25%
Daten-Management	80%	-	20%
Auswahl veröffentlichungswürdiger Daten	100%	-	-
Statistische Analyse	90%	-	10%
Erstellung von Tabellen und Abbildungen	100%	-	-
Manuskript Management bis zur Veröffentlichung	80%	10%	10%

## 1.2 Publikation: "The dentist's reliability to detect dental pathologies on periapical radiographs"

	Theresa Meusburger	Prof. Dr. Kühnisch	Co-Autoren (Annika Wülk)
Projektidee/Studiendesign	25%	50%	25%
Literaturrecherche	100%	-	-
Bildselektion	100%	-	-
Training und Kalibrierung der Arbeitsgruppe	25%	50%	25%
Daten-Management	80%	-	20%
Auswahl veröffentlichungswürdiger Daten	100%	-	-
Statistische Analyse	90%	-	10%
Erstellung von Tabellen und Abbildungen	100%	-	-
Manuskript Management bis zur Veröffentlichung	80%	10%	10%

## 2. Einleitung

Bildgebende Verfahren sind im zahnärztlichen Alltag ein routinemäßig verwendetes, diagnostisches Hilfsmittel in Ergänzung zur klinischen Untersuchung (Fuhrmann, 2013; Mallya und Lam, 2019; Shah et al., 2014). Ein aktuelles Röntgenbild, das den Zustand der zu behandelnden Zähne zeigt, ist eine wichtige Grundlage für eine indikationsgerechte Therapieentscheidung. Es ermöglicht Aussagen über restaurative Maßnahmen, z.B. Versorgung des Zahnes mit einer direkten oder indirekten Restauration oder Wurzelfüllung (WF), pathologische Zustände der Zahnhartsubstanzen, z.B. Karies, oder des Zahnhalteapparates, z.B. apikale Läsionen, Parodontitis oder Paro-Endo-Läsionen. Ebenso erlaubt es dem Zahnarzt das vorhandene Knochenangebot für eine Versorgung mit einem Implantat zu beurteilen und besser einschätzen zu können, ob dafür ein Knochenaufbau notwendig wird. Dabei handelt es sich um zahnärztliche Befunde, welche typischerweise mit der visuellen Inspektion nicht eindeutig beurteilbar sind (Fuhrmann, 2013; Mallya und Lam, 2019; Shah, 2014). Da Karies, apikale Entzündungen und parodontale Erkrankungen aber bis heute zu den häufigsten Erkrankungen der Mundhöhle zählen (GBD et al., 2020; Kassebaum et al., 2017; Tibúrcio-Machado et al., 2021), stellt die Versorgung mit indirekten und direkten Restaurationen, Wurzelfüllungen und bei Zahnverlust mit Implantaten eine alltägliche Aufgabe für einen Zahnarzt dar. Aus diesem Grund ist eine zuverlässige Erkennung und Beurteilung dieser restaurativen Behandlungsmaßnahmen auf Röntgenbildern unerlässlich. Allerdings unterliegt die Diagnostik aller medizinischen wie auch zahnmedizinischen Befunde der subjektiven Kompetenz eines jeden Untersuchers und abweichende Befunde sind daher keine Seltenheit. Dies begründet sich mit einem unterschiedlichen Wissenserwerb oder divergierenden klinischen Erfahrungen. Da bislang nur eine begrenzte Anzahl an Studien vorliegt

(Auderset et al., 2022; Mauad et al., 2021; Eckerbom et al., 1986; Saunders et al., 2000; Sebring et al., 2021), welche sich mit der korrekten Analyse der Inter- und Intra-Untersucher-Reliabilität von restaurativen Behandlungsmaßnahmen auf apikalen Zahnfilmen auseinandersetzen, erscheinen weitere Studien dazu notwendig. Die Erkennung von Restaurationen wurde in anderen Publikationen vorwiegend auf Einschätzung ihrer Suffizienz (Mauad et al., 2021) oder in anderen Bereichen der Medizin wie der Forensik zur zuverlässigen Ermittlung des Zahnstatus untersucht (Auderset et al., 2022). Auch andere Diagnostikmethoden wie intraorale Fotografien kamen vermehrt zum Einsatz (Mesinger et al., 2022). Zu Wurzelfüllungen fanden sich einzelne Vergleichswerke (Eckerbom et al., 1986; Saunders et al., 2000; Sebring et al., 2021), da hier neben der Orthopantomographie (OPG) der apikale Zahnfilm die Diagnostikmethode der Wahl darstellt. Wird der gegenwärtige Wissensstand zusammengefasst, so wird allerdings deutlich, dass der Fokus von Zuverlässigkeitsstudien mit dem Ziel die Inter- und Intra-Untersucher-Reliabilität zu ermitteln, in der Vergangenheit vor allem auf der radiologischen Befundung von Karies im Zusammenhang mit der Entwicklung neuer Diagnostikmethoden lag (u.a. Neuhaus et al., 2022; Kühnisch et al., 2021; Janjic Rankovic et al., 2021; Kapor et al., 2021; Schwendicke et al., 2015; Twetmann et al., 2013; Litzemberger et al., 2018). Die begrenzte Datenlage, insbesondere im Hinblick auf die Verwendung von restaurativen Behandlungsmaßnahmen als Diagnostikkategorien und apikalen Zahnfilmen als Diagnostikmethode, unterstreicht einmal mehr die Bedeutung dieser Zuverlässigkeitsstudie und hebt sich mit der Bewertung mehrerer Kategorien in einer Untersuchung als Alleinstellungsmerkmal von anderen Reliabilitätsstudien ab.

### **3. Zielstellung**

Das Ziel des vorliegenden Dissertationsprojektes bestand in der Untersuchung der Inter- und Intra-Untersucher-Zuverlässigkeit bei der Diagnostik von ausgewählten zahnärztlichen Restaurationsmaßnahmen (direkte Restaurationen/indirekte Restaurationen/Wurzelfüllungen/Implantate) auf apikalen Zahnfilmen mit Hilfe der Ermittlung von Cohens Kappa (CK)-Werten. Zudem sollte die Hypothese überprüft werden, dass die Variablen „Diagnostikkategorie“, „Evaluierungsrunde“ und „Berufserfahrung des Untersuchers“ einen statistisch messbaren Einfluss auf die Ergebnisse zeigen.

## 4. Material und Methoden

Die vorliegende Untersuchung ist Teil des Studienprojektes „Diagnostik mit Künstlicher Intelligenz (KI)“ an der Poliklinik für Zahnerhaltung und Parodontologie der LMU München (Projektnummer 20-0798) und zielte darauf ab, Stärken und Schwächen in der röntgenologischen Diagnostik von Zahnärzten zu erkennen, um erforderliche Kalibrierungen zielorientiert umzusetzen.

Das zu beurteilende Set (N=150) an apikalen Zahnfilmen bestand aus vollständig anonymisierten, digitalen Röntgenbildern, die anhand bestimmter Kriterien im Vorfeld aus einem großen Bilderpool (N>500) ausgewählt worden sind. Vor Beginn der Evaluierung durchliefen alle Untersucher ein theoretisches und praktisches Training unter Anleitung von Prof. Dr. Kühnisch.

Die vier ausgewählten Diagnostikkategorien und deren diagnostischen Standards, auf dessen Grundlage jeder Untersucher seine Wertung abgeben sollte, wurden im Vorfeld gemäß den aktuellen Richtlinien festgelegt. Die Kategorie Restaurationen wurde in direkte und indirekte Restaurationen unterschieden und unabhängig von ihrer Lage, Form und Röntgenopazität analysiert. Zu direkten Restaurationen zählten Komposit-, Zement- und Amalgamfüllungen. Keramik-, Verblend-Metall-Keramik (VMK)- oder Metall- Voll- und Teilkronen wurden der Kategorie indirekte Restaurationen zugeordnet. Eine weitere Kategorie stellten Wurzelfüllungen dar, die unabhängig von ihrer Suffizienz (Länge der WF, vollständige und homogene WF, insuffiziente WF durch fehlende Randständigkeit oder Luftblasen) analysiert worden sind (Ng et al., 2008). Des Weiteren bildeten Implantate verschiedener Hersteller eine eigene Diagnostikkategorie. Trotz des Einschlusses von verschiedenen Restaurationsmaterialien und Implantatsysteme-

men in das Bilderset, erfolgte in dieser Studie keine gesonderte Beurteilung dieser Merkmale. Ebenso wurde die Suffizienz der restaurativen Maßnahmen (Hickel et al., 2007, 2010, 2022) nicht bewertet.

Der ausgewählte Bildersatz von 150 apikalen Zahnfilmen wurde daraufhin von 14 Zahnärzten (davon drei mit mehr als zwei Jahren und 11 mit weniger als zwei Jahren Berufserfahrung) in zwei voneinander unabhängigen Evaluierungsrunden mit einem zeitlichen Abstand von mindestens vier Wochen beurteilt. Für jede Diagnostikkategorie wurde eine dichotome Entscheidung (ja/nein) getroffen. Im Anschluss an die zweite Evaluierungsrunde wurde der Referenzstandard mit allen beteiligten Zahnärzten erarbeitet bis ein Konsens erzielt wurde. Die statistische Auswertung der Daten erfolgte mittels Excel und SPSS (SPSS Statistics 27, 2020, IBM Corporation, Armonk, NY, USA). Zur Bestimmung der Inter- und Intra-Reliabilität der Untersucher untereinander und in Relation zum Referenzstandard wurden CK-Werte und deren Mittelwerte berechnet. Alle CK-Werte wurden gemäß den internationalen Standards nach Cohen (1960), Fleiss und Cohen (1973) und Landis und Koch (1977) interpretiert und ausgewertet. Im weiteren Verlauf wurde eine binäre logistische Regressionsanalyse unter Verwendung des Rückwärtseliminationsmodells für das Ergebnis (richtige/falsche Diagnoseentscheidung in Bezug auf den Referenzstandard) und für die Variablen „Diagnostikkategorie“, „Evaluierungsrunde“ und „Berufserfahrung des Zahnarztes“ erstellt.

## 5. Ergebnisse

Für alle ausgewählten restaurativen Maßnahmen wurden CK-Mittelwerte der Inter-Untersucher-Reliabilität in einem exzellenten Bereich (0.844-0.994) für beide Evaluierungsrunden dokumentiert. Den höchsten CK-Wert erzielten Implantate im ersten Durchgang (0.994). Die CK-Mittelwerte für Wurzelkanalfüllungen (0.920) und indirekten Restaurationen (0.932) waren in der ersten Runde nahezu identisch, gefolgt von direkten Restaurationen mit einem geringfügig niedrigeren Kappa-Wert (0.859). Die Intra-CK-Werte bewegten sich in der gleichen Größenordnung (0.910-0.987). In Relation zum Referenzstandard zeichneten sich ebenso nahezu perfekte Inter-CK-Werte ab (0.837-0.996), die nur minimale Abweichungen zwischen den einzelnen Kategorien aufzeigten. Das logistische Regressionsmodell bestätigte, dass Implantate am zuverlässigsten (OR 1) und direkte Restaurationen (OR 0.03) am wenigsten zuverlässig bewertet wurden und ein signifikanter Unterschied zwischen den Diagnostikkriterien vorliegt (p-Wert <0.001). Die Variablen „Evaluierungsrunde“ und „Berufserfahrung“ zeigten dagegen in der vorliegenden Studie keine Signifikanz (p-Wert 0.266, 0.755).

**Tabelle 1.** Berechnung der Odds Ratios (OR) mit entsprechenden 95%-Konfidenzintervallen (CI) und p-Werten auf Grundlage des logistischen Regressionsmodells

Variablen	Gruppen	OR	95% CI	p-Werte
<i>Diagnostikkategorie</i>	Implantate	1	-	-
	Direkte Restaurationen	0.03	0.02-0.06	<0.001
	Indirekte Restaurationen	0.08	0.04-0.17	<0.001
	Wurzelfüllungen	0.21	0.10-0.43	<0.001
<i>Evaluierungsrunde</i>	1	1	-	-
	2	0.90	0.74-1.09	0.266
<i>Berufserfahrung</i>	<2 Jahre	1	-	-
	>2 Jahre	1.04	0.82-1.31	0.755

## 6. Diskussion

Das vorliegende Dissertationsprojekt verfolgte das Ziel, die Inter- und Intra-Untersucher-Reliabilität für die Erkennung von restaurativen Behandlungsmaßnahmen auf apikalen Zahnfilmen zu bestimmen und zu vergleichen. Insgesamt wurden für alle vier Diagnostikkategorien nahezu perfekte CK-Werte aufgezeichnet. Die aufgestellte Hypothese, dass kein Unterschied zwischen den einzelnen Diagnostikkategorien vorliegt, musste verworfen werden. Dies konnte mit den ermittelten CK-Werten und der Auswertung des logistischen Regressionsmodells widerlegt werden. Dagegen zeigten die Variablen „Berufserfahrung“ und „Evaluierungsrunde“, anders als vor Durchführung der Studie angenommen, keinen statistisch messbaren Einfluss auf die Ergebnisse. Betrachtet man die Abweichungen zwischen den einzelnen Diagnostikkategorien im Detail, sollten die Unterschiede in der Erkennung restaurativer Maßnahmen nicht überbewertet werden. Implantate, Metallkronen oder Amalgamrestorationen sind große, röntgenopake Strukturen (Okuda et al., 2010), die auf apikalen Zahnfilmen deutlicher erkennbar sind, als kleine Kompositfüllungen oder grazile, temporäre Wurzelfüllungen, die nur eine geringe Röntgenopazität aufweisen (Tagger und Katz, 2004). Dies könnte neben Dokumentationsfehlern die Ursache für die geringfügig niedrigeren Ergebnisse in diesen Kategorien gewesen sein.

Durchaus relevant für einen Zahnarzt ist auch die zuverlässige Erkennung pathologischer Befunde auf Röntgenbildern. Aus diesem Grund wurde eine weitere Studie von Meusburger et al. (2023) durchgeführt, die den Bezug zur vorliegenden Untersuchung herstellen sollte. Hier wurde die Erkennung von Karies, apikale Läsionen, parodontalem Knochenabbau (PBL) und Paro-Endo-Läsionen auf apikalen Röntgenbildern analysiert. Im Gegensatz zur Studie von Wülk et al. (2023), in der alle CK-Werte durchweg in einem nahezu perfekten Bereich lagen,

erwies sich die Beurteilung von pathologischen Befunden, insbesondere für Zahnärzte mit wenig Berufserfahrung, als herausfordernd. Dies wurde durch Kappa-Werte in einem nur moderaten bis substanziellen Bereich beschrieben (Meusburger et al., 2023). Es ist jedoch zu erwähnen, dass es für einen unerfahrenen Zahnarzt womöglich schwieriger ist einen pathologischen Befund einheitlich zu diagnostizieren als eine Restauration zu erkennen. Eine forensische Studie von Auderset et al. (2022) ermittelte dagegen vergleichbare CK-Werte für die Erkennung von direkten Restaurationen; jedoch zeigten sich auch hier erhebliche Unterschiede zwischen erfahrenen und unerfahrenen Zahnärzten. Vergleiche mit weiteren Studien waren nur begrenzt möglich, da bislang der Fokus in der Erkennung von restaurativen Maßnahmen überwiegend auf der Beurteilung der Suffizienz, wie z.B. der Kronenpassung, lag (Mauad et al., 2021). Für Wurzelfüllungen sind dagegen vergleichbare Publikationen verfügbar. Hier wurden in den Studien von Saunders et al. (2000) und Sebring et al. (2021) ähnliche (substanziell bis nahezu perfekte) Werte für die Erkennung von Wurzelfüllungen beschrieben; in der Vergangenheit wurden aber auch moderate CK-Werte dokumentiert (Eckerbom et al., 1986). Untersuchungen, die sich mit der Erkennung von Implantaten beschäftigten, wurden bislang nicht publiziert. Ein weiterer interessanter Punkt war die Tatsache, dass im Vergleich zu der Veröffentlichung von Sebring et al. (2021) in den Studien von Wülk et al. (2023) und Meusburger et al. (2023) keine Verbesserung von der ersten zur zweiten Evaluierungsrunde erfolgte. Dies könnte wiederum auf das Training zurückzuführen sein, da aus diesem Grund alle Untersucher (auch diejenigen mit wenig Berufserfahrung) schon vor Durchführung der Studie einen guten Wissenstand aufwiesen.

Obwohl das vorliegende Studienprojekt insgesamt ein breites Spektrum an alltäglichen Diagnosen auf Röntgenbildern abdeckte, ist der Faktor, dass die klinische Situation nicht beurteilt wurde, als Schwäche anzusehen. Diese ist im zahnärztlichen Alltag eine wichtige Grundlage für die richtige Diagnosestellung (Fuhrmann, 2013; Mallya und Lam, 2019; Shah et al., 2014). Des Weiteren sind einige Problematiken in Bezug auf das Studiendesign zu nennen. Von jedem Untersucher wurde nur eine dichotome Entscheidung getroffen ohne die Suffizienz der restaurativen Maßnahmen beurteilen zu müssen, was die Entscheidung erleichtert und damit die Werte positiv beeinflusst haben könnte. Ebenso muss betont werden, dass nur apikale Zahnfilme als Diagnostikmethode verwendet wurden. Daher sind die Ergebnisse nicht vollständig auf andere bildgebende Verfahren wie Panoramaschicht- oder Bissflügelröntgenaufnahmen übertragbar. Die Verwendung anderer Diagnostikmethoden oder die Befundung restaurativer Maßnahmen im Zusammenhang mit der Erkennung pathologischer Befunde könnte ein aufschlussreicher Ansatz für zukünftige Studienprojekte sein. Zudem sollte berücksichtigt werden, dass im analysierten Bildersatz Röntgenbilder mit Mängeln ausgeschlossen worden sind, obwohl diese im klinischen Alltag ebenso zuverlässig ausgewertet werden müssen. Ein weiterer Schwachpunkt der Studie war einerseits die ungleiche Verteilung der Diagnostikkategorien in den ausgewählten Röntgenbildern; andererseits war die Untersucherguppe nicht gleichmäßig in unerfahrene und erfahrene Untersucher aufgeteilt. Zuletzt muss erwähnt werden, dass zur Bildbetrachtung von den Untersuchern keine standardisierten Monitore verwendet wurden, was sich ebenfalls auf die radiologische Diagnose ausgewirkt haben könnte (Moshfeghi et al., 2015). Abschließend muss betont werden, dass die aufgeführten Schwächen auch zu Verzerrungen in den Ergebnissen geführt haben könnten.

## 7. Zusammenfassung und Ausblick

Zusammengefasst lassen die Ergebnisse aus der vorliegenden Studie darauf schließen, dass die Erkennung der untersuchten restaurativen Behandlungsmaßnahmen für erfahrene und unerfahrene Zahnärzte als souverän und zuverlässig eingestuft werden kann. Demgegenüber stehen die Ergebnisse aus der Untersuchung von Meusburger et al. (2023), die zeigen, dass für die Erkennung von pathologischen Befunden, insbesondere von apikalen Läsionen und PBL, ein erhöhter Schulungsbedarf besteht. Dies ist eine interessante Erkenntnis, der primär von Universitäten Beachtung geschenkt werden sollte. Dementsprechend sollte der Fokus von künftigen Lehr- und Fortbildungsprogrammen auf den genannten Schwachstellen in der radiologischen Diagnostik liegen. Auch eine Untersuchung, in der ein genaueres Scoring in Bezug auf die Suffizienz restaurativer Maßnahmen oder das Ausmaß pathologischer Befunde durchgeführt wird, wäre ein interessanter Ansatz für zukünftige Forschungsvorhaben, um weitere Unsicherheiten in der Befundung von Röntgenbildern zu erkennen.

Darüber hinaus ist es im Hinblick auf die rapide Entwicklung von Methoden der KI in der Medizin und Zahnmedizin erforderlich, mögliche Schwachstellen in der röntgenologischen Diagnostik von Ärzten und Zahnärzten, die an der Entwicklung dieser Systeme beteiligt sind, zu identifizieren. Folglich können frühzeitig und zielorientiert Kalibrierungen der Studiengruppen umgesetzt und damit langfristig ein allgemeingültiger und reproduzierbarer Qualitätsstandard sichergestellt werden. Mit Blick in die Zukunft könnten diese KI-basierten Systeme dann auch als Unterstützung bei der Befundung von Röntgenbildern und damit bei einer automatisierten und standardisierten Diagnosefindung dienen (Schwendicke et al., 2020).

## 8. Englische Zusammenfassung

The present dissertation project aimed to determine and compare the inter- and intra-examiner reliability for the detection of restorative measures on periapical radiographs. For the detection of restorative measures, overall almost perfect Cohens Kappa values were recorded indicating a high reliability for experienced and inexperienced dentists. In contrast, the assessment of pathological findings proved challenging, especially for dentists with little professional experience, as evidenced in the investigation by Meusburger et al. (2023) by CK values in a moderate to substantial range. Apical inflammation and PBL had the lowest CK values and were thus rated least reliable whereas the category implants achieved the highest CK value overall. The hypothesis that there is no difference between the individual diagnostic categories had to be rejected. This could be occupied with the determined Kappa values and the logistic regression model. The variable "evaluation round" appeared to be not significant in both studies. In contrast, the results for the variable "dentist's clinical experience" differed. In the present study, no significant difference was found between the reliability values of the examiners, whereas in the research by Meusburger et al. (2023), experienced dentists achieved better CK values. In conclusion, these results show that the focus of training programs should be on the aforementioned weak points, especially the detection of apical lesions and PBL. Also, a study in which a more precise scoring is performed with regard to the sufficiency of restorative measures or the extent of pathological findings would be an interesting approach for future research projects in order to identify further uncertainties in the reporting of radiographic diagnostics. In view of the rapid development of artificial intelligence (AI) methods in medicine and dentistry, it is necessary to identify possible weaknesses in the radiological diagnostics of physicians and dentists involved in the

development of these systems. Consequently, calibrations of the study groups can be implemented early and in a target-oriented manner, thus ensuring a generally valid and reproducible quality standard in the long term. Looking to the future, these AI-based systems could serve as support for radiological diagnostics and thus also for automated and standardized diagnosis finding (Schwendicke et al., 2020).

## 9. Veröffentlichung I

Wülk, A., Meusburger, T., Dujic, H., Hickel, R., Kessler, A., Heck, K., Kühnisch, J. (2023). The Dentist's Ability to Detect Different Restorative Procedures on Periapical Radiographs-Results from a Reliability Study. *International Journal of Environmental Research and Public Health*, 20(3), 2619. <https://doi.org/10.3390/ijerph20032619>

## 10. Veröffentlichung II

Meusburger, T., Wülk, A., Kessler, A., Heck, K., Hickel, R., Dujic, H., Kühnisch, J. (2023). The Detection of Dental Pathologies on Periapical Radiographs- Results from a Reliability Study. *Journal of Clinical Medicine*, 12(6), 2224. <https://doi.org/10.3390/jcm12062224>

## 11. Literaturverzeichnis

- Auderset, F. C., Connert, T., Meller, C., Filippi, A., Dagassan-Berndt, D. C. (2022). Evaluation of five methods to identify composite restorations in human teeth on a forensic purpose-an ex vivo comparative study. *International journal of legal medicine*. Online ahead of print.
- Cohen, J. A. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and physiological measurement*, 20, 37–46.
- Eckerbom, M., Andersson, J. E., Magnusson, T. (1986). Interobserver variation in radiographic examination of endodontic variables. *Endodontics and dental traumatology*, 2(6), 243–246.
- Fleiss, J. L., Cohen, J. (1973). The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability. *Educational and physiological measurement*, 33, 613–619.
- Fuhrmann, A. (2013). *Zahnärztliche Radiologie* (1. Auflage). Stuttgart: Thieme.
- GBD 2017 Oral Disorders Collaborators, Bernabe, E., Marcenes, W., Hernandez, C. R., Bailey, J., Abreu, L. G., Alipour, V., Amini, S., Arabloo, J., Arefi, Z., Arora, A., Ayanore, M. A., Bärnighausen, T. W., Bijani, A., Cho, D. Y., Chu, D. T., Crowe, C. S., Demoz, G. T., Demsie, D. G., Dibaji Forooshani, Z. S., Kassebaum, N. J. (2020). Global, Regional, and National Levels and Trends in Burden of Oral Conditions from 1990 to 2017: A Systematic Analysis for the Global Burden of Disease 2017 Study. *Journal of dental research*, 99(4), 362–373.
- Hickel, R., Roulet, J. F., Bayne, S., Heintze, S. D., Mjör, I. A., Peters, M., Rousson, V., Randall, R., Schmalz, G., Tyas, M., Vanherle, G. (2007). Recommendations for conducting controlled clinical studies of dental restorative materials. *Clinical oral investigations*, 11(1), 5–33.
- Hickel, R., Peschke, A., Tyas, M., Mjör, I., Bayne, S., Peters, M., Hiller, K. A., Randall, R., Vanherle, G., Heintze, S. D. (2010). FDI World Dental Federation: clinical criteria for the evaluation of direct and indirect restorations-update and clinical examples. *Clinical oral investigations*, 14(4), 349–366.
- Hickel, R., Mesinger, S., Opdam, N., Loomans, B., Frankenberger, R., Cadenaro, M., Burgess, J., Peschke, A., Heintze, S. D., Kühnisch, J. (2022). Revised FDI criteria for evaluating direct and indirect dental restorations-recommendations for its clinical use, interpretation, and reporting. *Clinical oral investigations*. Online ahead of print.

- Janjic Rankovic, M., Kapor, S., Khazaei, Y., Crispin, A., Schüler, I., Krause, F., Ekstrand, K., Michou, S., Eggmann, F., Lussi, A., Huysmans, M. C., Neuhaus, K., Kühnisch, J. (2021). Systematic review and meta-analysis of diagnostic studies of proximal surface caries. *Clinical oral investigations*, 25(11), 6069–6079.
- Kapor, S., Rankovic, M. J., Khazaei, Y., Crispin, A., Schüler, I., Krause, F., Lussi, A., Neuhaus, K., Eggmann, F., Michou, S., Ekstrand, K., Huysmans, M. C., Kühnisch, J. (2021). Systematic review and meta-analysis of diagnostic methods for occlusal surface caries. *Clinical oral investigations*, 25(8), 4801–4815.
- Kassebaum, N. J., Smith, A. G. C., Bernabé, E., Fleming, T. D., Reynolds, A. E., Vos, T., Murray, C. J. L., Marcenes, W., GBD 2015 Oral Health Collaborators (2017). Global, Regional, and National Prevalence, Incidence, and Disability-Adjusted Life Years for Oral Conditions for 195 Countries, 1990-2015: A Systematic Analysis for the Global Burden of Diseases, Injuries, and Risk Factors. *Journal of dental research*, 96(4), 380–387.
- Kühnisch, J., Janjic Rankovic, M., Kapor, S., Schüler, I., Krause, F., Michou, S., Ekstrand, K., Eggmann, F., Neuhaus, K. W., Lussi, A., Huysmans, M. C. (2021). Identifying and Avoiding Risk of Bias in Caries Diagnostic Studies. *Journal of clinical medicine*, 10(15), 3223.
- Landis, J. R., Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Litzenburger, F., Heck, K., Pitchika, V., Neuhaus, K. W., Jost, F. N., Hickel, R., Jablonski-Momeni, A., Welk, A., Lederer, A., Kühnisch, J. (2018). Inter- and intraexaminer reliability of bitewing radiography and near-infrared light transillumination for proximal caries detection and assessment. *Dento maxillo facial radiology*, 47(3), 20170292.
- Mallya, S. M., Lam, E. W. N. (2019). *White and pharoah's oral radiology: principles and interpretation* (8th edition). St. Louis: Elsevier.
- Mauad, L. Q., Doriguêto, P. V. T., Almeida, D., Fardim, K. A. C., Machado, A. H., Devito, K. L. (2021). Quantitative assessment of artefacts and identification of gaps in prosthetic crowns: a comparative in vitro study between periapical radiography and CBCT images. *Dento maxillo facial radiology*, 50(3), 20200134.
- Mesinger, S., Heck, K., Crispin, A., Frankenberger, R., Cadenaro, M., Burgess, J., Peschke, A., Heintze, S. D., Loomans, B., Opdam, N., Hickel, R., Kühnisch, J. (2022). Evaluation of direct restorations using the revised FDI criteria: results from a reliability study. *Clinical oral investigations*. Online ahead of print.

- Meusburger, T., Wülk, A., Kessler, A., Heck, K., Hickel, R., Dujic, H., Kühnisch, J. (2023). The Detection of Dental Pathologies on Periapical Radiographs-Results from a Reliability Study. *Journal of clinical medicine*, 12(6), 2224.
- Moshfeghi, M., Shahbazian, M., Sajadi, S. S., Sajadi, S., Ansari, H. (2015). Effects of Different Viewing Conditions on Radiographic Interpretation. *Journal of dentistry* (Tehran, Iran), 12(11), 853–858.
- Neuhaus, K. W., Eggmann, F., Kühnisch, J., Kapor, S., Janjic Rankovic, M., Schüler, I., Krause, F., Lussi, A., Michou, S., Ekstrand, K., Huysmans, M. C. (2022). STandard Reporting of CAries Detection and Diagnostic Studies (STARCARDDS). *Clinical oral investigations*, 26(2), 1947–1955.
- Ng, Y. L., Mann, V., Rahbaran, S., Lewsey, J., Gulabivala, K. (2008). Outcome of primary root canal treatment: systematic review of the literature -- Part 2. Influence of clinical factors. *International endodontic journal*, 41(1), 6–31.
- Okuda, Y., Noda, M., Kono, H., Miyamoto, M., Sato, H., Ban, S. (2010). Radio-opacity of core materials for all-ceramic restorations. *Dental materials journal*, 29(1), 35–40.
- Saunders, M. B., Gulabivala, K., Holt, R., Kahan, R. S. (2000). Reliability of radiographic observations recorded on a proforma measured using inter- and intra-observer variation: a preliminary study. *International endodontic journal*, 33(3), 272–278.
- Sebring, D., Kvist, T., Buhlin, K., Jonasson, P., EndoReCo, Lund, H. (2021). Calibration improves observer reliability in detecting periapical pathology on panoramic radiographs. *Acta odontologica Scandinavica*, 79(7), 554–561.
- Schwendicke, F., Tzschope, M., Paris, S. (2015). Radiographic caries detection: A systematic review and meta-analysis. *Journal of dentistry*, 43(8), 924–933.
- Schwendicke, F., Samek, W., Krois, J. (2020). Artificial Intelligence in Dentistry: Chances and Challenges. *Journal of dental research*, 99(7), 769–774.
- Shah, N., Bansal, N., Logani, A. (2014). Recent advances in imaging technologies in dentistry. *World journal of radiology*, 6(10), 794–807.
- Tagger, M., Katz, A. (2004). A standard for radiopacity of root-end (retrograde) filling materials is urgently needed. *International endodontic journal*, 37(4), 260–264.
- Tibúrcio-Machado, C. S., Michelon, C., Zanatta, F. B., Gomes, M. S., Marin, J. A., Bier, C. A. (2021). The global prevalence of apical periodontitis: a systematic review and meta-analysis. *International endodontic journal*, 54(5), 712–735.

- Twetman, S., Axelsson, S., Dahlén, G., Espelid, I., Mejäre, I., Norlund, A., Tranæus, S. (2013). Adjunct methods for caries detection: a systematic review of literature. *Acta odontologica Scandinavica*, 71(3-4), 388–397.
- Wülk, A., Meusburger, T., Dujic, H., Hickel, R., Kessler, A., Heck, K., Kühnisch, J. (2023). The Dentist's Ability to Detect Different Restorative Procedures on Periapical Radiographs-Results from a Reliability Study. *International journal of environmental research and public health*, 20(3), 2619.

## **Danksagung**

An dieser Stelle bedanke mich von ganzem Herzen bei allen, die mich während der gesamten Zeit dieses Dissertationsprojektes unterstützt haben und ohne deren Mithilfe, die Anfertigung dieser Arbeit so nicht möglich gewesen wäre:

Besonderer Dank gilt meinem Doktorvater, Herrn Prof. Dr. Jan Kühnisch, für seine ausgezeichnete Betreuung, die tatkräftige, fachliche Unterstützung, sein außerordentliches Engagement und die durchweg angenehme Arbeitsatmosphäre.

Des Weiteren danke ich meinen Promotionskolleginnen und Co-Autoren, Helena Dujic und Theresa Meusburger, für den fachlichen Austausch und die freundschaftliche, unerlässliche Unterstützung. Auch allen anderen Mitarbeitern und Doktoranden der Arbeitsgruppe danke ich für die schöne Zeit und die angenehme, kollegiale Zusammenarbeit.

Zuletzt gebührt meinen Eltern ein großer Dank, die mir das Studium der Zahnmedizin und die Promotion ermöglicht haben und mir jederzeit mit Rat und Tat zur Seite standen. Ebenso danke ich meiner Familie und Freunden für das immer offene Ohr und die stete, moralische Unterstützung und Motivation.