

Aus der

Universitätsklinik Tübingen

Abteilung Diagnostische und interventionelle Radiologie

**Bewertung der diagnostischen Güte der intraoperativen
Präparatetomosynthese im Vergleich zur Präparate-
radiographie bei Mammakarzinompatientinnen**

**Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Medizin**

**der Medizinischen Fakultät
der Eberhard Karls Universität
zu Tübingen**

**vorgelegt von
Walder, Ramona (geb. Schnell)**

2024

Dekan: Professor Dr. B. Pichler

1. Berichterstatter: Privatdozentin Dr. H. Preibsch

2. Berichterstatter: Privatdozentin Dr. C. Bachmann

Tag der Disputation: 21.03.2024

Inhaltsverzeichnis

Inhaltsverzeichnis	I
Abbildungsverzeichnis	V
Tabellenverzeichnis	X
Abkürzungsverzeichnis	XII
1 Einleitung	1
1.1 Epidemiologie des Mammakarzinoms	1
1.2 Histopathologische Klassifikation des Mammakarzinoms	2
1.3 Präoperative Brustkrebsdiagnostik	4
1.3.1 2D-Mammographie	4
1.3.2 Tomosynthese	6
1.3.3 Sonographie	7
1.3.4 Kontrastmittel-Magnetresonanz-Tomographie	8
1.4 Operative Therapie des Mammakarzinoms	8
1.5 Intraoperative Präparatediagnostik	10
1.6 Fragestellung und Arbeitshypothese	11
2 Material und Methoden	13
2.1 Herkunft Datenmaterial	13
2.2 Verwendete Geräte	14
2.3 Histopathologische Auswertung	15
2.4 Bewertungsbögen	16
2.5 Verwendete mathematische Methoden	16
2.5.1 Konkordanzanalyse	16
2.5.2 Verwendete Software	18
3 Ergebnisse	19

3.1	Studienpopulation und -kollektiv	19
3.2	Ergebnisse präoperative Mammographie und Inter-Reader-Vergleich	21
3.3	Ergebnisse intraoperative PR und Inter-Reader-Vergleich	27
3.3.1	Präparatgröße.....	28
3.3.2	Größe Herdbefund	29
3.3.3	Randabstand.....	31
3.3.4	Inter-Reader-Vergleich der Diagnose R0/R1	33
3.4	Ergebnisse intraoperative DBT und Inter-Reader-Vergleich	34
3.4.1	Präparatgröße.....	34
3.4.2	Größe Herdbefund	36
3.4.3	Randabstand.....	37
3.4.4	Inter-Reader-Vergleich der Diagnose R0/R1	39
3.5	Vergleich der PR und DBT	40
3.5.1	Vergleich von PR und DBT anhand aggregierter Ergebnisgrößen des Inter-Reader-Vergleichs	40
3.5.2	Inter-Methoden Vergleich der Diagnose R0/R1	42
3.6	Vergleich des Goldstandards (Histologie) mit PR und DBT	44
3.6.1	Vergleich der maximalen Herdbefundgröße	45
3.6.2	Vergleich der Randabstandsmessung	49
3.6.3	Vergleich der R0/R1-Diagnose	52
3.6.4	Sensitivität des R0/R1-Vergleichs auf Randabstandskriterien	58
3.6.5	Vergleich der Identifikation von Mikrokalk.....	61
3.6.6	Zeitlicher Aufwand zur Analyse der Präparate in den radiologischen Verfahren	66
3.7	Zusammenfassung der Ergebnisse.....	67
4	Diskussion.....	70

4.1	Vergleich zum Stand der Wissenschaften.....	70
4.2	Kritische Diskussion der Ergebnisse und Einflussfaktoren.....	76
5	Zusammenfassung und Ausblick	79
6	Literatur.....	81
7	Erklärung zum Eigenanteil	93
8	Anhang.....	94
8.1	Bewertungsbogen Mammographie	94
8.2	Bewertungsbogen intraoperative PR.....	95
8.3	Bewertungsbogen intraoperative DBT	96
8.4	Weitere Diagramme Inter-Reader-Vergleich Mammographie	97
8.5	Weitere Diagramme Inter-Reader-Vergleich PR	100
8.5.1	Präparatgröße.....	100
8.5.2	Größe Herdbefund	102
8.5.3	Randabstand.....	104
8.6	Weitere Diagramme Inter-Reader-Vergleich DBT	109
8.6.1	Präparatgröße.....	109
8.6.2	Größe Herdbefund	111
8.6.3	Randabstand.....	113
8.7	Vergleich zum Goldstandard – Weitere Diagramme und Kontingenztafeln	118
8.7.1	Größe Herdbefund (Reader 2).....	118
8.7.2	Randabstand (Reader 2)	120
8.7.3	Kontingenztafeln R0/R1-Vergleich (Reader 2).....	122
8.7.4	Sensitivität des R0/R1-Vergleichs auf Randabstandskriterien (Reader 2).....	126
8.7.5	Sensitivität des R0/R1-Vergleichs auf Randabstandskriterien für originale Diagnosen der Histologie (Reader 1 und 2)	128

8.7.6	Kontingenztafeln Erkennbarkeit Mikrokalk (Reader 2).....	131
8.8	Abhängigkeit der Güte der R0/R1-Diagnose von der Herdbefundgröße	133

Abbildungsverzeichnis

Abbildung 2.1: Bildung der Datensätze zur weiteren Untersuchung	14
Abbildung 3.1: Histogramm Altersverteilung (Datensatz 1).....	19
Abbildung 3.2: Anteile nach Entität in der Stichprobe	20
Abbildung 3.3: Anteil der Brustdichte in der Stichprobe (Datensatz 1)	22
Abbildung 3.4: Vergleich der Entität aus Mammographie mit der Histologie (Datensatz 1)	23
Abbildung 3.5: Inter-Reader-Vergleich der Größe des Herdbefundes am Beispiel CCx (MG)	24
Abbildung 3.6: Anteil MK in der Stichprobe (MG)	26
Abbildung 3.7: Inter-Reader-Vergleich der Größe der MK am Beispiel MLOy (MG).....	26
Abbildung 3.8: Inter-Reader-Vergleich der Größe des Präparats am Beispiel APx (PR).....	28
Abbildung 3.9: Inter-Reader-Vergleich der Größe des Herdbefundes am Beispiel APx (PR)	30
Abbildung 3.10: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel APkaudal (PR).....	32
Abbildung 3.11: Inter-Reader-Vergleich der Größe des Präparats am Beispiel APx (DBT).....	35
Abbildung 3.12: Inter-Reader-Vergleich der Größe des Herdbefundes am Beispiel APx (DBT)	36
Abbildung 3.13: Inter-Reader-Vergleich des Randabstandes am Beispiel APkaudal (DBT)	38
Abbildung 3.14: Subjektive Einschätzung der Reader auf die Frage wie gut der Herdbefund in DBT und PR messbar ist	42
Abbildung 3.15: Herausforderungen beim Vergleich der Ergebnisse aus der Histologie und den radiologischen Verfahren.....	44
Abbildung 3.16: Vergleich der maximalen Größe des Herdbefundes zwischen PR und Goldstandard (Reader 1)	46

Abbildung 3.17: Vergleich der maximalen Größe des Herdbefundes zwischen DBT und Goldstandard (Reader 1)	47
Abbildung 3.18: Vergleich des Randabstandes LLdorsal zwischen PR und Goldstandard (Reader 1)	49
Abbildung 3.19: Vergleich des Randabstandes LLdorsal zwischen DBT und Goldstandard (Reader 1)	51
Abbildung 3.20: Sensitivitätsanalyse der R0/R1-Diagnose der PR auf Randabstandskriterien (Reader 1)	59
Abbildung 3.21: Sensitivitätsanalyse der R0/R1-Diagnose der DBT auf Randabstandskriterien (Reader 1)	60
Abbildung 3.22: Subjektive Einschätzung der Reader auf die Frage wie gut MK in DBT und PR erkennbar und messbar sind	65
Abbildung 3.23: Subjektive Einschätzung der Reader auf die Frage wie der zeitliche Aufwand der Messdatenerhebungen in DBT und PR empfunden wurde	66
Abbildung 8.1: Bewertungsbogen Mammographie	94
Abbildung 8.2: Bewertungsbogen intraoperative PR	95
Abbildung 8.3: Bewertungsbogen intraoperative DBT	96
Abbildung 8.4: Inter-Reader Vergleich der Größe des Herdbefundes am Beispiel CCz (MG)	97
Abbildung 8.5: Inter-Reader-Vergleich der Größe des Herdbefundes am Beispiel MLOy (MG)	97
Abbildung 8.6: Inter-Reader-Vergleich der Größe des Herdbefundes am Beispiel MLOz (MG)	98
Abbildung 8.7: Inter-Reader-Vergleich der Größe der MK am Beispiel CCx (MG)	98
Abbildung 8.8: Inter-Reader-Vergleich der Größe der MK am Beispiel CCy (MG)	99
Abbildung 8.9: Inter-Reader-Vergleich der Größe der MK am Beispiel MLOz (MG)	99
Abbildung 8.10: Inter-Reader-Vergleich der Größe des Präparats am Beispiel APx (PR)	100

Abbildung 8.11: Inter-Reader-Vergleich der Größe des Präparats am Beispiel LLy (PR).....	101
Abbildung 8.12: Inter-Reader-Vergleich der Größe des Präparats am Beispiel LLz (PR).....	101
Abbildung 8.13: Inter-Reader-Vergleich der Größe des Herdbefundes am Beispiel APz (PR)	102
Abbildung 8.14: Inter-Reader-Vergleich der Größe des Herdbefundes am Beispiel LLz (PR)	103
Abbildung 8.15: Inter-Reader-Vergleich der Größe des Herdbefundes am Beispiel LLy (PR)	103
Abbildung 8.16: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel APkranial (PR).....	104
Abbildung 8.17: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel APlateral (PR).....	105
Abbildung 8.18: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel APmammilär (PR).....	105
Abbildung 8.19: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel APmedial (PR).....	106
Abbildung 8.20: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel APperipher (PR)	106
Abbildung 8.21: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel LLdorsal (PR)	107
Abbildung 8.22: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel LLkutan (PR)	107
Abbildung 8.23: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel LLmammilär (PR)	108
Abbildung 8.24: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel LLperipher (PR).....	108
Abbildung 8.25: Inter-Reader-Vergleich der Größe des Präparats am Beispiel APx (DBT).....	109
Abbildung 8.26: Inter-Reader-Vergleich der Größe des Präparats am Beispiel LLy (DBT).....	110

Abbildung 8.27: Inter-Reader-Vergleich der Größe des Präparats am Beispiel LLz (DBT).....	110
Abbildung 8.28: Inter-Reader-Vergleich der Größe des Herdbefundes am Beispiel APz (DBT)	111
Abbildung 8.29: Inter-Reader-Vergleich der Größe des Herdbefundes am Beispiel LLy (DBT)	112
Abbildung 8.30: Inter-Reader-Vergleich der Größe des Herdbefundes am Beispiel LLz (DBT)	112
Abbildung 8.31: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel APkranial (DBT).....	113
Abbildung 8.32: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel APlateral (DBT)	114
Abbildung 8.33: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel APmammilär (DBT)	114
Abbildung 8.34: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel APmedial (DBT).....	115
Abbildung 8.35: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel APperipher (DBT)	115
Abbildung 8.36: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel LLdorsal (DBT)	116
Abbildung 8.37: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel LLkutan (DBT)	116
Abbildung 8.38: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel LLmammilär (DBT)	117
Abbildung 8.39: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel LLperipher (DBT).....	117
Abbildung 8.40: Vergleich der maximalen Größe des Herdbefundes zwischen PR und Goldstandard (Reader 2)	118
Abbildung 8.41: Vergleich der maximalen Größe des Herdbefundes zwischen DBT und Goldstandard (Reader 2)	119
Abbildung 8.42: Vergleich des Randabstandes LLdorsal zwischen PR und Goldstandard (Reader 2)	120

Abbildung 8.43: Vergleich des Randabstandes LLdorsal zwischen DBT und Goldstandard (Reader 2)	121
Abbildung 8.44: Sensitivitätsanalyse der R0/R1-Diagnose der PR auf Randabstandskriterien (Reader 2)	126
Abbildung 8.45: Sensitivitätsanalyse der R0/R1-Diagnose der DBT auf Randabstandskriterien (Reader 2)	127
Abbildung 8.46: Sensitivitätsanalyse der R0/R1-Diagnose der PR auf Randabstandskriterien (Reader 1/ Originale R0/R1-Diagnose) ..	128
Abbildung 8.47: Sensitivitätsanalyse der R0/R1-Diagnose der PR auf Randabstandskriterien (Reader 2/ Originale R0/R1-Diagnose) ..	129
Abbildung 8.48: Sensitivitätsanalyse der R0/R1-Diagnose der DBT auf Randabstandskriterien (Reader 1/ Originale R0/R1-Diagnose) ..	129
Abbildung 8.49: Sensitivitätsanalyse der R0/R1-Diagnose der DBT auf Randabstandskriterien (Reader 2/ Originale R0/R1-Diagnose) ..	130
Abbildung 8.50: Relative Übereinstimmung der R0/R1-Diagnose aus der PR für Fälle innerhalb der jeweiligen Intervalle der Herdbefundgrößen (Reader 1).....	133
Abbildung 8.51: Relative Übereinstimmung der R0/R1-Diagnose aus der PR für Fälle innerhalb der jeweiligen Intervalle der Herdbefundgrößen (Reader 2).....	134
Abbildung 8.52: Relative Übereinstimmung der R0/R1-Diagnose aus der DBT für Fälle innerhalb der jeweiligen Intervalle der Herdbefundgrößen (Reader 1).....	134
Abbildung 8.53: Relative Übereinstimmung der R0/R1-Diagnose aus der DBT für Fälle innerhalb der jeweiligen Intervalle der Herdbefundgrößen (Reader 2).....	135

Tabellenverzeichnis

Tabelle 1.1: Einteilung der Brustdichte	5
Tabelle 2.1: Beurteilung der Übereinstimmung anhand von Cohens Kappa .	18
Tabelle 3.1: Vergleich Anteile nach Altersgruppen mit Daten aus dem Krebsregister Baden-Württemberg	20
Tabelle 3.2: Vergleich der Verteilung der Entität mit Literaturdaten.....	21
Tabelle 3.3: Vergleich der Verteilung der Brustdichte mit Literaturdaten	22
Tabelle 3.4: Inter-Reader-Vergleichswerte Größe des Herdbefundes (MG)..	25
Tabelle 3.5: Inter-Reader-Vergleichswerte Größe MK (MG).....	27
Tabelle 3.6: Inter-Reader-Vergleichswerte Größe des Präparats (PR).....	29
Tabelle 3.7: Inter-Reader-Vergleichswerte Größe des Herdbefundes (PR)...	30
Tabelle 3.8: Inter-Reader-Vergleichswerte Größe des Randabstandes (PR)	32
Tabelle 3.9: Kontingenztafel R0/R1-Vergleich intraoperative PR	33
Tabelle 3.10: Inter-Reader-Vergleichswerte Größe des Präparats (DBT)	36
Tabelle 3.11: Inter-Reader-Vergleichswerte Größe des Herdbefundes (DBT)	37
Tabelle 3.12: Inter-Reader-Vergleichswerte Größe des Randabstandes (DBT)	39
Tabelle 3.13: Kontingenztafel R0/R1-Vergleich intraoperative DBT	39
Tabelle 3.14: Gemittelte Vergleichswerte zur Beurteilung der Inter-Reader Reliabilität der intraoperativen DBT und PR.....	41
Tabelle 3.15: Kontingenztafel zur Inter-Methoden-Reliabilität R0/R1 zwischen intra-operativer PR und DBT (Reader 1).....	43
Tabelle 3.16: Kontingenztafel zur Inter-Methoden-Reliabilität R0/R1 zwischen intra-operativer PR und DBT (Reader 2).....	43
Tabelle 3.17: Vergleichswerte für die max. Größe des Herdbefundes für PR und DBT im Vergleich zur Histologie	48
Tabelle 3.18: Vergleich der Randabstandsmessung von Reader 1 und Reader 2 mit der Histologie (PR).....	51
Tabelle 3.19: Vergleich der Randabstandsmessung von Reader 1 und Reader 2 mit der Histologie (DBT).....	52
Tabelle 3.20: Kontingenztafel R0/R1-Vergleich PR und Histologie (Reader 1)	53

Tabelle 3.21: Kontingenztafel R0/R1-Vergleich PR und Histologie (Reader 1, Datensatz 2).....	54
Tabelle 3.22: Kontingenztafel R0/R1-Vergleich DBT und Histologie (Reader 1)	56
Tabelle 3.23: Kontingenztafel R0/R1-Vergleich DBT und Histologie (Reader 1, Datensatz 2).....	57
Tabelle 3.24: Vergleich der Sensitivität und Spezifität zwischen den radiologischen Verfahren und der Histologie	58
Tabelle 3.25: Kontingenztafel zum Vergleich der Erkennbarkeit von MK in der PR und der Histologie (Reader 1)	62
Tabelle 3.26: Kontingenztafel zum Vergleich der Erkennbarkeit von MK in der DBT und der Histologie (Reader 1).....	63
Tabelle 3.27: Vergleich der Sensitivität und Spezifität von PR und DBT bei der Erkennbarkeit von MK.....	64
Tabelle 8.1: Kontingenztafel R0/R1-Vergleich PR und Histologie (Reader 2)	122
Tabelle 8.2: Kontingenztafel R0/R1-Vergleich PR und Histologie (Reader 2, Datensatz 2).....	123
Tabelle 8.3: Kontingenztafel R0/R1-Vergleich DBT und Histologie (Reader 2)	124
Tabelle 8.4: Kontingenztafel R0/R1-Vergleich DBT und Histologie (Reader 2, Datensatz 2).....	125
Tabelle 8.5: Kontingenztafel zum Vergleich der Erkennbarkeit von MK in der PR und der Histologie (Reader 2)	131
Tabelle 8.6: Kontingenztafel zum Vergleich der Erkennbarkeit von MK in der DBT und der Histologie (Reader 2).....	132

Abkürzungsverzeichnis

Abkürzungen

2D	2-dimensional
3D	3-dimensional
Abb.	Abbildung
ACR	American College of Radiology
AP	Anterior-posterior
BET	Brusterhaltende Therapie
BI-RADS	Breast imaging reporting and data system
CC	Craniocaudal
CIS	Carcinoma in situ
DBT	Digitale Brust-Tomosynthese
DCIS	Duktales carcinoma in situ
ER	Estrogenrezeptor
Gy	Gray
HER2	Human epidermal growth factor receptor 2
ILC	Invasiv lobuläres carcinoma
K2	Cohens Kappa
KM-MRT	Kontrastmittel-Magnetresonanz-Tomographie
LCIS	Lobuläres carcinoma in situ
LL	Laterolateral
LOA	Limits-of-Agreement
MG	Mammographie
MK	Mikrokalzifikationen
MLO	Mediolateral-oblique
mm	Millimeter
MRT	Magnetresonanz-Tomographie
N	Newton
n	Number
NST	No specific type (Mammakarzinomentität)

OP	Operation
PACS	Picture archiving and communication System
PR	Präparateradiographie
PgR	Progesteronrezeptor
SEN	Sensitivität
SPE	Spezifität
TNM	Tumor-Node-Metastasis (Tumorklassifikationssystem)
UICC	Union for international cancer control

Das in dieser Arbeit gewählte generische Femininum bezieht sich zugleich auf männliche, weibliche und diverse Geschlechteridentitäten. Alle Geschlechteridentitäten werden ausdrücklich mitgemeint, soweit die Aussagen dies erfordern.

1 Einleitung

1.1 Epidemiologie des Mammakarzinoms

In der westlichen Welt ist das Mammakarzinom die häufigste maligne Erkrankung der Frau [1, 2]. Bei steigender Inzidenz von annähernd 70.000 jährlichen Fällen in Deutschland (Stand 2018) erkrankt ungefähr jede achte Frau (12,4 %) im Laufe ihres Lebens an Brustkrebs [3]. Die wachsende Inzidenz lässt sich teilweise durch das zwischen 2005 und 2009 eingeführte landesweite Mammographie-Screening und die damit verbundene verbesserte Früherkennung erklären. Zusätzlich kann die Ausbreitung gewisser Risikofaktoren in der weiblichen Bevölkerung, wie beispielsweise Adipositas, bestimmte reproduktive Faktoren (Nulliparität, höheres Alter der Erstgebärenden) und die wachsende allgemeine Lebenserwartung, ebenfalls die Inzidenz erhöhen [4–8]. Trotz steigender Neuerkrankungen ist die brustkrebsbedingte Mortalitätsrate seit Anfang der 1990-er Jahre, durch verbesserte Therapiemöglichkeiten und die zunehmende Detektion des Tumors bereits in präinvasiven oder frühen invasiven Stadien, rückläufig [1, 3, 8, 9]. Auf Basis der Daten von 2018 liegt das Lebenszeitrisiko an Brustkrebs zu sterben bei 3,5 % (1 von 29 erkrankten Frauen) [3]. Das Mammakarzinom bleibt, bezogen auf alle Krebserkrankungen ungeachtet der sinkenden Mortalität die häufigste Todesursache der Frau [9, 10] und ist bei Frauen zwischen 35 und 55 Jahren die häufigste Todesursache überhaupt [11]. Durch die hohe Prävalenz in unserer Gesellschaft ist eine optimale Primärversorgung der Patientin ausschlaggebend, insbesondere um das Auftreten von Rezidiven und die daraus resultierenden individuellen und wirtschaftlichen Folgen bestmöglich zu vermeiden [1, 12, 13]. Das Auftreten eines Lokalrezidivs innerhalb eines frühen Zeitraums (<12 Monate) beeinflusst die Prognose der Patientin negativ und senkt die 5-Jahresüberlebensrate im Vergleich zu Patientinnen die kein Lokalrezidiv entwickeln [14]. Darüber hinaus führt ein Rezidiv und die damit verbundene erneute Therapie zu weiteren Einschränkungen in der Lebensqualität der Patientin wie beispielsweise die Zunahme von

kognitiven Beeinträchtigungen [15] sowie das erhöhte Auftreten von Depressionen, Angststörungen und Anpassungsstörungen [16, 17].

1.2 Histopathologische Klassifikation des Mammakarzinoms

Die Mehrheit aller Mammakarzinome entwickelt sich entweder aus den duktalem Epithelzellen (duktale Karzinome) oder aus den Drüsenläppchen (lobuläre Karzinome) und entsteht in rund 50 % der Fälle, aufgrund des dort vorhandenen größten Drüsengewebs-zu-Brustvolumenverhältnisses, im oberen äußeren Quadranten [5, 11]. 20 % befinden sich im Mamillenbereich, jeweils 10 % in den restlichen Quadranten [5]. Die größte Gruppe (60-80 %) der invasiven Karzinome wird durch ihr heterogenes Erscheinungsbild keinem morphologisch speziellen Typ (tubulär, medullär, papillär, muzinös) zugeordnet und folglich als no specific type (NST) kategorisiert [1, 2, 5, 18]. Invasiv lobuläre Karzinome (ILC) bilden mit 10-15 % die zweithäufigste Mammakarzinom-Entität [5].

Neben invasiven Karzinomen gibt es auch nicht-invasive Tumore, die carcinoma in situ (CIS), welche sich auf die Milchgänge bzw. auf die Drüsenläppchen beschränken [1, 5]. Charakteristisch ist hier eine neoplastische Proliferation bereits maligner Zellen innerhalb der Drüsenschläuche bei intakter Basalmembran. Eine Metastasierung ist aufgrund der intakten Basalmembran bei reinen in-situ Läsionen nicht möglich [1, 5]. Anhand des speziellen Wachstumsmusters können bei den Vorläuferläsionen das duktales carcinoma in situ (DCIS) und das lobuläre carcinoma in situ (LCIS) unterschieden werden, welche 95 % bzw. 5 % der CIS ausmachen [2, 5].

Beide präinvasive Formen können mit radiologisch sichtbaren Mikrokalzifikationen (MK) assoziiert sein, wodurch eine präsymptomatische Detektion möglich ist [18]. Im Jahr 2017 waren 20 % der im Mammographie-Screening nachgewiesenen Läsionen ein DCIS, was als ein Erfolg für die Früherkennung interpretiert werden darf [19].

Das Risiko, dass Patientinnen mit einem DCIS ein invasives Karzinom entwickeln, das hämatogen und lymphogen metastasieren kann, ist mehr als doppelt so hoch verglichen mit der Normalbevölkerung [20]. Zudem zeigt die postoperative Histologie bei Frauen, die ursprünglich wegen eines DCIS operiert wurden,

in 20 % der Fälle ein zusätzliches invasives Karzinom [21] und bei mehr als 80 % der operierten invasiven Karzinome wird histopathologisch ein zusätzliches DCIS beschrieben [18, 22]. Diese Prognose führt dazu, dass bei DCIS ebenso wie bei einem invasiven Karzinom die operative Entfernung des Tumors empfohlen wird [8].

Die postoperative histopathologische Aufarbeitung eines eingesandten Mammapräparats umfasst neben der Klassifikation des Karzinomtyps (z.B. invasiv lobuläres Karzinom) die Stadieneinteilung in das TNM-System, den Malignitätsgrad gemäß Elston und Ellis sowie die Angabe eines möglichen Residualtumors [23]. Mit Hilfe der TNM-Klassifikation werden Tumore anhand ihres maximalen Größendurchmessers (T), des regionären Nodalbefalls (N) und der Fernmetastasierung (M) eingruppiert [23]. Das Stadium nach der Union for international cancer control (UICC) wird aus dieser Klassifikation abgeleitet. Anhand der Tubulusausbildung, der Kernpolymorphie und der Mitoserate wird der Malignitätsgrad gemäß Elston und Ellis bestimmt [5]. Hierbei wird ein Summenscore der genannten Kriterien gebildet und folglich eine Gruppierung in die unterschiedlichen Malignitätsgrade G1 (gut differenziert), G2 (mäßig differenziert) oder G3 (schlecht differenziert) vorgenommen [5]. Ein Residualtumor liegt vor, wenn makroskopisch (R2) oder mikroskopisch (R1) maligne Zellen im Körper der Patientin oder am Rand des operierten Präparats nachweisbar sind. Ein R0-Status (keine malignen Zellen nachweisbar) ist das primäre Ziel der operativen Entfernung des Karzinoms [1, 5, 12].

Von besonderer Bedeutung ist hierbei, dass die Lage der gemessenen Resektionsränder korrekt zugeordnet wird, um im Falle einer Re-Operation das Nachresektat gezielt entnehmen zu können [11, 24]. Neben den aufgeführten histologischen Klassifikationen hängt die individuelle Prognose der Patientin von dem ermittelten Rezeptorstatus (ER, PgR, HER2) und ihrem Alter ab [8]. Der vorliegende Karzinomtyp, das Grading und der Rezeptorstatus werden bereits präoperativ mit Hilfe der Stanz- oder Vakuumbiopsie des in der Bildgebung auffälligen Befundes bestimmt. Die erhobenen Tumorcharakteristika werden zusammen mit den klinischen und bildgebenden Befunden in einer

interdisziplinären Tumorkonferenz diskutiert und die Therapieoptionen anschließend mit der Patientin besprochen [1].

Eine optimale Therapie (Vermeidung von Über,- und Untertherapie) und das damit angestrebte rezidivfreie Überleben kann nur auf Basis einer korrekten histologischen Diagnose und unter Berücksichtigung von multiplen individuellen Faktoren erfolgen [11, 12].

1.3 Präoperative Brustkrebsdiagnostik

In Deutschland werden Frauen im Alter von 50-69 Jahren zweijährlich zum Mammographie-Screening eingeladen [19]. Ziel dieser sekundären Prävention ist die Detektion eines möglichen Karzinoms im präinvasiven oder frühen invasiven Stadium [1, 10]. Es wird somit das übergeordnete Ziel einer Mortalitätssenkung verfolgt [2, 8, 19]. Die landesweite standardisierte Empfehlung für die Diagnostik, Therapie und Nachsorge des Mammakarzinoms wird in der Stufe-3-Leitlinie beschrieben. Neben der klassischen 2D-Mammographie können zusätzliche bildgebende Verfahren wie Tomosynthese, Sonographie und MRT bei der Einschätzung eines möglichen Befundes hilfreich sein [8].

1.3.1 2D-Mammographie

Die Mammographie ist nach aktuellem medizinischen Kenntnisstand die einzige allgemein anerkannte Methode für die Brustkrebsfrüherkennung [1, 2, 8, 11]. Hierbei stellt die digitale Vollfeldmammographie in zwei Ebenen, mediolateral-oblique (MLO) und craniocaudal (CC), das bildgebende Standardverfahren dar [1, 25]. Zur Reduktion von Streustrahlung sowie von Bewegungsartefakten und der damit verbundenen Dosisreduktion wird die Aufnahme der Brust in Kompression durchgeführt [26]. Zur standardisierten Beschreibung eines Herdbefundes, der eine dreidimensionale, raumfordernde Veränderung darstellt, gehören Form, Randbegrenzung und Dichte [27]. Spezielle Merkmale einer Läsion, beispielsweise kleine Verkalkungen (MK), können bereits einen Hinweis auf eine mögliche Malignität geben [2, 28, 29].

30-60 % der invasiven Karzinome [2, 30] und 90 % der präinvasiven Vorstufen (meist DCIS) sind mit mammographisch detektierbaren Kalzifikationen assoziiert

[1, 2, 11, 30]. Durch die radiologisch erkennbaren Verkalkungen ist die Sensitivität der Mammographie selbst in der dichten Brust hoch [2, 11, 30].

Die Einschätzung der Brustdichte ist Teil der standardisierten Interpretation und Befundung der Mammographie, welche durch die international anerkannte Klassifikation Breast Imaging Reporting and Data System (BI-RADS) des American College of Radiology (ACR) eingeteilt wird [27]. Die neuste (fünfte) Auflage von 2013 kategorisiert die Brustdichte mithilfe der Buchstaben A-D (siehe Tabelle 1.1) statt wie in der vorherigen Version mit Zahlen (I-IV), um eine mögliche Verwechslung mit der Einteilung in die BI-RADS- Klassifikation zur Befundkategorisierung zu vermeiden [27]. Zudem wird auf eine prozentuale Angabe der Brustdichte verzichtet und stattdessen eine sensitivitätsbezogene Einteilung gefordert [28].

Tabelle 1.1: Einteilung der Brustdichte

Brustdichte ACR	Beschreibung
A	die Brust ist überwiegend lipomatös (Involutionsbrust)
B	es finden sich vereinzelt fibroglanduläre Verdichtungen (Teilinvolution)
C	inhomogen dichtes Brustgewebe (weitgehend dichter Drüsenkörper)
D	extrem dichtes Mammaparenchym (sehr dichter Drüsenkörper)

Eine dichte Brust ist ein unabhängiger Risikofaktor für das Auftreten von Brustkrebs [8] und erschwert zusätzlich die Interpretation bildgebender Diagnostik [1]. Die annähernd einhundertprozentige Sensitivität der Mammographie bei vollständiger Fettgewebsinvolution (ACR A) [31], nimmt bei steigender Brustdichte durch mögliche Drüsenparenchymüberlagerungen kontinuierlich ab [29, 31, 32] und kann in der extrem dichten Brust auf Werte knapp unter 50 % sinken [33]. Hieraus resultiert, dass insgesamt 10-20 % [1, 30] der Karzinome und bei ACR D Brustdichte mehr als 39 % der Karzinome mammographisch nicht erkannt werden [34].

Durch die erschwerte Interpretation kann es neben der Nichterkennung möglicher Karzinome bei dichter Brust auch zu einer Überbehandlung der Patientin bei einem vermeintlich karzinomverdächtigen Befund kommen [29], der sich später als falsch-positiv darstellt [2]. Allgemein ist eine höhere Strahlendosis (3-5 fach) nötig, um eine dichte Brust (ACR C und ACR D) mammographisch zu untersuchen [32]. Dies sollte bei der Entscheidung der optimalen bildgebenden Diagnostik mit einbezogen werden. Zurzeit empfiehlt die S3-Leitlinie als präoperative Standardbildgebung bei histologisch gesichertem Karzinom die Mammographie plus Sonographie [8]. Ein Mamma-MRT kann hilfreich sein und wird unter bestimmten Indikationen (z.B. erhöhtes familiäres Risiko) durchgeführt [8]. Die Tomosynthese (siehe Abschnitt 1.3.2) wird derzeit als Zusatzaufnahme eingesetzt und hat ihren Stellenwert bislang in der Optimierung der mammographischen Diagnostik [8] sowie in der Biopsiesteuerung [35].

Ein ergänzendes Verfahren zusätzlich zur Mammographie kann nicht nur in der Früherkennung in Problembereichen (dichte Brust, familiäres Risiko) helfen, sondern sollte auch präoperativ zur bestmöglichen Einschätzung des Tumorausmaßes eingesetzt werden [8]. In einer Studie [36] wurden 42 % der Patientinnen mit ACR D (ehemals IV) aufgrund eines Tumors im Schnittrand erneut operiert. Eine dichte Brust ist folglich mit einem höheren R1-Risiko assoziiert [36].

1.3.2 Tomosynthese

Die digitale Brust-Tomosynthese (DBT) ist ein Verfahren, bei dem die rotierende Röntgenröhre eine Bildserie der Brust erstellt, so dass das Brustvolumen anschließend in Einzelschichten betrachtet werden kann [25, 26, 30]. Die Information ermöglicht eine umfassende Einschätzung des strukturellen Brustgewebes, insbesondere bei hoher Dichte [37], indem Parenchymüberlagerungen reduziert und Herdbefunde besser detektiert werden können [30, 38]. Diese gesteigerte Detektionsrate von malignen Herdbefunden konnte in unterschiedlichen Studien gezeigt werden [39–42]. Folglich lässt sich festhalten, dass die Sensitivität durch den Einsatz der DBT gesteigert wird [8]. Insbesondere unerfahrene Radiologinnen profitieren von der ergänzenden Bildgebung [43]. Auch die

Spezifität steigt durch die Kombination von Mammographie und DBT und ermöglicht eine höhere diagnostische Sicherheit [44, 45], was verschiedene Studien durch eine Reduktion der Wiedereinbestellungsrate um 17-57 % zeigen konnten [39, 40, 46, 47]. Eine Verbesserung der Spezifität wirkt sich positiv auf mögliche individuelle und wirtschaftliche Folgen, wie eine Reduktion der psychischen Belastung der Patientin und die Kostenreduktion durch den Wegfall von überflüssiger zusätzlicher Diagnostik, aus [1, 47].

Um eine zusätzliche Strahlenbelastung der Patientin zu vermeiden kann aus dem Datensatz der DBT eine synthetische 2D-Mammographie gewonnen werden [25]. Diese synthetische Mammographie ist dem ursprünglich durchgeführten combo mode (Tomosynthese + digitale Mammographie) nicht unterlegen [48, 49]. Neben der präzisen Beurteilung von Herdbefunden inklusive Messung der Tumorgroße [50, 51] kann die DBT auch während der brusterhaltenden Operation zur Präparatebildgebung insbesondere bei der Messung des Randabstandes und der daraus resultierenden Einschätzung R0/R1 hilfreich sein [52]. Einige Studien konnten bereits zeigen, dass der intraoperative Einsatz der DBT im Vergleich zur Präparateradiographie, welches das derzeitige 2D-Standardverfahren darstellt, die Re-Exzisionsrate senkt und folglich die Sensitivität gesteigert werden kann [53–55]. Trotz längerer erforderlicher Auswertungszeit [45, 56] wird das Bildmaterial der DBT im allgemeinen von Radiologinnen bevorzugt [37, 47].

1.3.3 Sonographie

Zur empfohlenen Basisuntersuchung bei auffälligem Befund oder im Rahmen der prätherapeutischen Diagnostik bei gesichertem Mammakarzinom gehört neben der Anamnese, der klinischen Tastuntersuchung und der Mammographie die Mammasonographie [8, 26, 30]. Die sonographische Untersuchung umfasst neben der gesamten Brustdrüse zusätzlich die Axilla [8, 26, 30]. Mithilfe der sonographisch gesteuerten Stanzbiopsie kann ein unklarer Befund histologisch abgeklärt werden [8]. Die Mammasonographie kann die Sensitivität der Mammographie als zusätzliche Diagnostik erhöhen, führt jedoch ebenfalls zu einer höheren falsch-positiven Rate und den damit verbundenen vermehrten

Kontrolluntersuchungen und Biopsien [8]. Zudem ist die Qualität der Sonographie Untersucherinnen-abhängig [57].

1.3.4 Kontrastmittel-Magnetresonanztomographie

Die Kontrastmittel-Magnetresonanztomographie (KM-MRT) ist die sensitivste bildgebende Diagnostik der Brust [30, 58] und hat bei Patientinnen mit familiärem Risiko als Früherkennungsmethode einen hohen Stellenwert [26]. Im Zuge dessen wird in den aktuellen S3-Leitlinien bei Hochrisiko-Patientinnen eine intensiviertere Früherkennung (iFE) empfohlen, die das MRT als wesentliche diagnostische Bildgebung miteinschließt [8]. Die MRT-Untersuchung und die darauffolgende radiologische Auswertung sind zeitlich aufwendig und mit hohen Kosten verbunden [58]. Demnach soll das MRT für die Brustkrebsdiagnostik bei Patientinnen nur eingesetzt werden, wenn die Mammographie plus Sonographie keine ausreichend eindeutige diagnostische Aussagekraft hat oder eine perkutane Biopsie aufgrund von unzureichender visueller Darstellung nicht möglich ist [8]. Zusätzlich ist die Spezifität im Vergleich zur hohen Sensitivität eingeschränkt, was bei einem Screening-MRT im Normalkollektiv weitere zahlreiche zeit- und kostenintensive Biopsien und Kontrolluntersuchungen bewirken würde [26].

1.4 Operative Therapie des Mammakarzinoms

Neben der Brustkrebsfrüherkennung ist die Verbesserung der Therapieoptionen eine wirksame Methode, um die brustkrebsbedingte Morbidität und Mortalität zu senken [1]. Die individuelle Therapie sollte in einem gemeinsamen Entscheidungsprozess zwischen behandelnder Ärztin und Patientin besprochen und gefunden werden. Hierbei sind das Tumorstadium inklusive der tumoreigenen biologischen Aktivität sowie das Alter und der Wunsch der Patientin zu berücksichtigen [12]. Unabhängig von einer möglichen neoadjuvanten/ adjuvanten Behandlung (Chemotherapie, Strahlentherapie, Hormontherapie) werden 87 % der Patientinnen mit Brustkrebs operiert [59]. Die operative Therapie mit R0-Resektion des Tumors ist das primäre Ziel. Gründe dafür keine Operation durchzuführen sind beispielsweise neben dem Patientinnen-Wunsch, weit fortgeschrittene Stadien, sowie eine hohe Morbidität [60]. 70 % der operativen

Tumorentfernungen werden in Deutschland als brusterhaltende Therapie (BET) und 30 % als Mastektomie durchgeführt [24, 59].

Indikation der BET sind lokal begrenzte, nicht-invasive Karzinome (DCIS) und invasive Karzinome, die mit einer günstigen Relation von Tumorgröße zu Brustvolumen einhergehen [1, 2, 8, 24].

Besteht bei einer Patientin die Möglichkeit zur brusterhaltenden Operation, wird der Tumor inklusive eines tumorfreien Randes entfernt. Eine intraoperative Clip-, oder Fadenmarkierung des Präparats ermöglicht später eine topographische Zuordnung [1, 24]. An die Operation anschließend wird eine Strahlentherapie empfohlen, um das Risiko der Entwicklung eines Lokalrezidivs zu minimieren [1, 2, 8, 24]. Bezogen auf das Überleben ist dieses Therapiekonzept genauso effektiv, um Brustkrebs zu behandeln, wie die Mastektomie [61] und überlegen in Bezug auf die Lebensqualität [62]. Eine Mastektomie sollte erwogen werden, wenn eine vollständige Exzision inklusive Nachresektion im Rahmen einer BET nicht möglich ist [24], eine Kontraindikation zur adjuvanten Therapie besteht, oder es dem Wunsch der aufgeklärten Patientin entspricht [8].

Unabhängig von der Operationstechnik wird am OP-Präparat beurteilt, ob der Tumor im Gesunden entfernt wurde. Das Ziel der kurativen, operativen Versorgung ist die komplette Exstirpation des Tumors inklusive tumorfreien Resektionsrändern (R0) [1, 8, 11, 24, 60]. Hierbei ist die R0-Klassifikation der Histologie bei invasiven Karzinomen nicht an einen allgemeingültigen, fest definierten metrischen Wert gebunden. Solange kein Tumor an der Schnittfläche nachweisbar ist, auch als „no ink on tumor“ bezeichnet, d.h. eine tumorfreie Gewebelamelle am Präparatrand nachweisbar ist, wird das Operationspräparat als R0 klassifiziert [24, 60, 63, 64]. Beim präinvasiven Karzinom (DCIS) hingegen wird bei vollständiger Exzision ein zusätzlicher Sicherheitsrand von mindestens 2 mm zwecks Reduktion des möglichen Rezidivrisikos angestrebt [8, 24, 65].

Generell minimieren tumorfreie Ränder das Rezidivrisiko, wobei größere Randabstände (>5 mm) keinen zusätzlichen Nutzen liefern [63].

Eine erneute Operation bei initial brusterhaltender Therapie wegen Tumornachweis am Resektionsrand (R1) kommt häufig (20-38 %) vor [66–68]. Besonders

hoch ist dieses Risiko für junge Patientinnen (dichte Brust) und Patientinnen mit ausgeprägtem DCIS [69].

Allgemein ist eine Re-Operation mit multiplen negativen Konsequenzen wie einem verzögerten Start der adjuvanten Therapie, einem schlechteren kosmetischen Ergebnis, höheren Kosten und einer größeren psychischen Belastung für die Patientin, verbunden [63, 66, 67].

1.5 Intraoperative Präparatediagnostik

Um bei der brusterhaltenden Therapie das Risiko einer Re-Operation zu minimieren, wird intraoperativ die Vollständigkeit des entnommenen Tumors mit Hilfe der Präparateradiographie (PR) beurteilt und dokumentiert [1].

Die PR ist hierbei das Standardverfahren und kann insbesondere bei nicht palpablen Tumoren wie DCIS, die präoperativ z.B. mit einem Drahtanker markiert wurden, hilfreich sein [70]. Neben der PR gibt es weitere mögliche Diagnostik wie beispielsweise intraoperativen Ultraschall, histologische Schnellschnittdiagnostik, Imprintzytologie und optische Spektroskopie (Gewebe-scanner), die unterschiedliche Vorzüge und Limitationen aufweisen [71–73]. In verschiedenen Studien [74–76] wurde die diagnostische Präzision (insb. R0/R1-Diagnose) der PR mit dem histologischen Goldstandard verglichen. Die Sensitivität der PR liegt, bezogen auf alle Tumorentitäten, bei 31-53 % und die Spezifität bei 75-84 % [74–76]. Insbesondere bei DCIS additional zur invasiven Komponente kann bei radiologisch imponierendem tumorfreiem Rand das tatsächliche histologische Ergebnis nur ungenügend vorhergesagt werden [75]. Hierbei wird die tatsächliche Größe der nicht-invasiven Komponente häufig unterschätzt [75, 77], was an fehlenden oder mammographisch nicht sichtbaren MK liegen könnte [78].

Bereits die Festlegung der Stelle, an der das Minimum des Randabstandes vorliegt und entsprechend zu messen ist, ist im Rahmen der radiologischen Befundung oft nicht standardisiert und wird nur in 55 % der Fälle korrekt identifiziert [75]. Der Grund für die Abweichung zwischen radiologischem und histologischem Befund könnten Projektionsphänomene, die Kompression des Präparats oder dessen postoperativer Volumenverlust sein [79, 80].

1.6 Fragestellung und Arbeitshypothese

Durch das Mammographie-Screening ist es möglich Brustkrebs bereits in den präinvasiven oder frühen invasiven Stadien zu detektieren, bevor eine klinische Symptomatik besteht [1]. Durch die meist sehr kleinen Primärtumore steigt die Indikation und Möglichkeit für eine BET [24]. Bereits 70 % der Frauen werden zurzeit brusterhaltend operiert, da bezogen auf die Mortalität die BET mit adjuvanter Bestrahlung gleichwertig zu der Mastektomie und überlegen bezogen auf das kosmetische Ergebnis und die allgemeine Lebensqualität der Patientin ist [24]. Das primäre Ziel ist es, den Tumor im Gesunden zu entfernen, also eine R0-Resektion zu gewährleisten. Eine R1-Resektion ist das größte Risiko für die Entwicklung eines Rezidivs [1, 11, 60, 63]. Die intraoperative Randbeurteilung ist hierfür ein entscheidungskritischer Zeitpunkt, deren Erfolg die zur Verfügung stehende Diagnostik mit hoher Sensitivität und Spezifität gewährleisten sollte [71–73]. Jede Fehlentscheidung kann für die Patientin eine weitere Operation oder ein Rezidiv und die damit verbundene erneute Therapie bedeuten [24, 63, 65]. Die PR ist derzeit das Standardverfahren, um intraoperativ die Resektionsränder zu beurteilen. Insbesondere bei nicht-palpablen Tumoren, die präoperativ mit einem Draht markiert wurden, kann die vollständige Exstirpation des Herdbefundes radiologisch nachvollzogen werden [1, 8]. Die PR ist ein schnelles und weit verbreitetes Verfahren, das kostengünstig eine objektive Momentaufnahme des Operationspräparats erstellt [1, 8].

Nachteilig sind die eingeschränkte Beurteilung von nicht verkalkten Läsionen, insbesondere bei dichtem Umgebungsgewebe und bei DCIS [78].

Verglichen mit dem späteren histologischen Befund weicht die PR, insbesondere bei der Einschätzung zum Randabstand und der folglich gestellten Diagnose R0/R1, in bisherigen Studien ab (vgl. Abschnitt 1.5) [81]. Die Steigerung der Sensitivität und Spezifität durch den zusätzlichen Einsatz einer DBT wurde präoperativ bereits in Studien bestätigt [39, 40, 42, 46].

Intraoperativ wurde der Einsatz der DBT ebenfalls in einigen Studien bewertet. Teils mit dem Ergebnis, dass die DBT keine bessere Einschätzung des histologischen Randabstandes ermöglicht [82]; teils mit dem Ergebnis, dass die Re-Exzisionsrate gesenkt und die Sensitivität somit gesteigert werden konnte [53–

55]. Auch Amer et al. stellte eine Überlegenheit der Tomosynthese im Gegensatz zur PR in Bezug auf die Diagnose R0/R1 verglichen mit der Histologie fest [83].

Die vorliegende Arbeit greift die Fragestellung der Güte von bildgebenden Verfahren zur Mammakarzinomdiagnostik auf und untersucht hierbei im Kern die nachfolgenden Punkte:

1. Evaluation der intraoperativen diagnostischen Wertigkeit der Tomosynthese (3D-Radiographie) im Vergleich zur 2D-Präparateradiographie in Bezug auf:
 - Robustheit des Verfahrens gegenüber unterschiedlichen Readern
 - Bewertung von Randabstand und R0/R1-Klassifikation
 - Bewertung von Mikrokalzifikationen

2. Schlussfolgerung auf die präoperative diagnostische Wertigkeit der Tomosynthese im Vergleich zur 2D-Mamographie auf Basis des vorliegenden Studienmaterials an Präparaten in Bezug auf:
 - Robustheit des Verfahrens gegenüber unterschiedlichen Readern
 - Bewertung der Größe des Herdbefundes
 - Bewertung von Mikrokalzifikationen

2 Material und Methoden

2.1 Herkunft Datenmaterial

Das im Rahmen dieser Arbeit verwendete Material stammt aus der randomisierten, prospektiven MAC-002 Studie, die vom 1. Quartal 2015 bis 2. Quartal 2017 an der Frauenklinik der Universitätsklinik Tübingen unter dem leitenden Prüfarzt Prof. Dr. med. Markus Hahn durchgeführt wurde. Primäres Studienziel der MAC-002 Studie war der Vergleich der Ultraschall-assistierten mit der konventionellen Tumorchirurgie bei Patientinnen mit Mammakarzinom und Indikation zur brusterhaltenden Therapie. Teil des sekundären Studienziels der MAC-002 Studie war die Beurteilung des intraoperativ mittels DBT und PR erstellten Bildmaterials der jeweiligen Operationspräparate. Dieses Ziel wird im Rahmen der vorliegenden Arbeit aufgegriffen und weiter ausgearbeitet.

Hierzu wurde das Bildmaterial von zwei Oberärztinnen der Abteilung für Interventionelle und Diagnostische Radiologie der Universitätsklinik Tübingen unter der Leitung des ärztlichen Direktors Herrn Prof. Dr. med. Konstantin Nikolaou, mit einer Berufserfahrung von 11 Jahren für Reader 1 und 10 Jahren für Reader 2, retrospektiv ausgewertet. Neben den intraoperativ erstellten PR- und DBT-Bildern wurden ebenfalls die zugehörigen präoperativen Mammographiebilder von beiden Readern beurteilt. Die Ergebnisse der radiologischen Auswertungen wurden mit den histopathologischen Befunden verglichen, die als Goldstandard gelten. Hierbei soll die Güte von bildgebenden Verfahren zur Mammakarzinomdiagnostik untersucht und deren diagnostische Wertigkeit evaluiert werden. Die vorliegende Studie ist durch die Ethikkommission der Universität Tübingen genehmigt (Nr. 128/2015BO1).

Im Rahmen der MAC-002 Studie wurden Patientinnen mit histologisch gesichertem Mammakarzinom und Indikation zur brusterhaltenden Operation eingeschlossen. Ausschlusskriterien waren ein sonographisch nicht darstellbarer Herdbefund, Multizentrität und eine neoadjuvante Vorbehandlung oder geplante neoadjuvante Therapie. In der vorliegenden Studie sind zudem Patientinnen mit reinem DCIS ausgeschlossen. Aus den ursprünglich 336 Patientinnen der MAC-002 Studie wurden 124 Patientinnen zufällig ausgewählt, bei

denen eine PR und eine DBT vorhanden ist, um das Kollektiv dieser Arbeit zu bilden. Mit den 124 Fällen wurden zwei verschiedene Datensätze gebildet, siehe Abbildung 2.1:

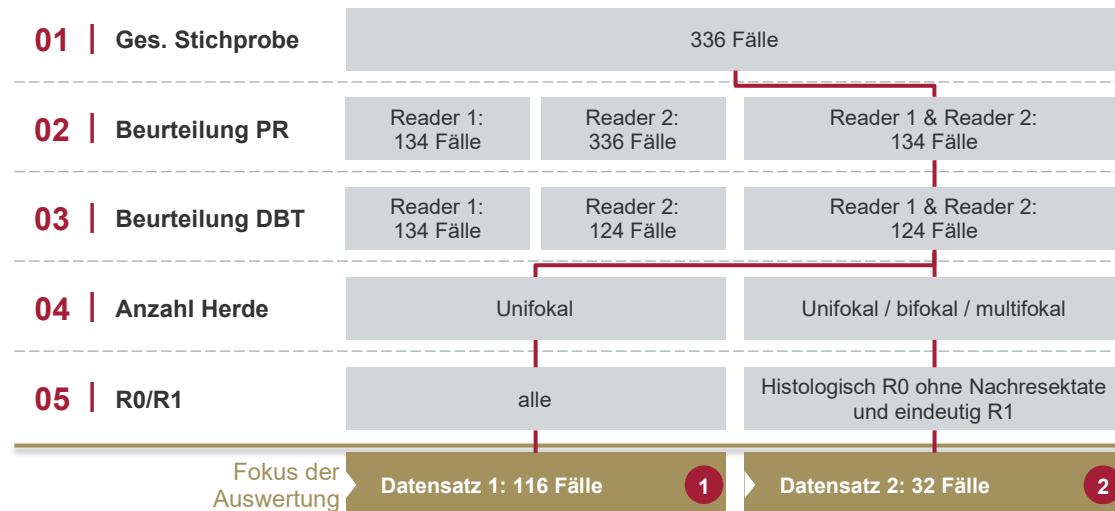


Abbildung 2.1: Bildung der Datensätze zur weiteren Untersuchung

In Datensatz 1 werden nur Patientinnen mit unifokalem Herd betrachtet. In Datensatz 2 werden alle Patientinnen eingeschlossen, bei denen die histologische Diagnose R0 bereits am Präparat ohne mögliche Nachresektate vorliegt und die Randabstände fast vollständig ≥ 10 mm betragen oder die Diagnose R1 vorliegt und der Befund so ausgeprägt ist, dass eine Infiltration des Tumors oder der nicht invasiven Komponente an mehreren Rändern festgestellt wird und folglich als „eindeutig“ eingestuft werden kann. Der zweite Datensatz soll beim Vergleich der radiologischen Methoden mit dem Goldstandard, der Histologie, mögliche subjektive Störeinflüsse der Reader durch Interpretationsspielraum in der Diagnose minimieren und wird zur Verifikation der Ergebnisse des Vergleichs aus Datensatz 1 verwendet.

2.2 Verwendete Geräte

Die bildgebenden Verfahren am Präparat (PR und DBT) wurden an einem Vollfeld-Mammographiesystem mit integrierter Tomosynthese-Funktion (Selenia Dimensions, Firma Hologic, USA) erstellt. Die präoperativen Mammographien sind teilweise ebenfalls an dem bereits genannten System generiert worden oder an einem GoldSeal Senographe Essential, teils kommt das Bildmaterial auch von

extern. Nach Erhalt des Operationspräparats wurde die 2D-Präparateradiographie standardgemäß in zwei Ebenen (anterior-posterior (AP) und laterolateral (LL)), teilweise unter Kompression, angefertigt. Anschließend ist mit einem 15 Grad Scanwinkel, unter kontinuierlicher Röhrenbewegung, die DBT durchgeführt worden. Die erstellten 25 Einzelbilder werden mittels mathematischer Algorithmen zu einem „quasi“ 3D-Bild rekonstruiert, wobei die Schichtdicke 1 mm beträgt. Teilweise wurden die DBT-Bilder ebenfalls unter Kompression generiert. Nach der diagnostischen Radiographie ist das Präparat zur histologischen Aufarbeitung an die Pathologie gesendet worden.

2.3 Histopathologische Auswertung

Die Operationspräparate inklusive intraoperativ gewonnener Nachresektate werden postoperativ makroskopisch und mikroskopisch vom Institut für Allgemeine Pathologie der Universitätsklinik Tübingen unter der Leitung des ärztlichen Direktors Herrn Prof. Dr. med. Falko Fend beurteilt. Die Diagnose bezüglich Entität, Präparatgröße, Tumorgröße und minimalem Abstand der Resektionsränder sowie der Aussage R0/R1, gelten als Goldstandard und wurden aus dem Bericht der postoperativen Histologie entnommen. Derzeit gilt eine invasive Komponente als in sano entfernt, wenn am Schnittrand kein Tumor nachweisbar ist („no ink on tumor“). Für die nicht invasive Komponente liegt der ausreichende Sicherheitsabstand bei ≥ 2 mm [8]. In der vorliegenden Studie wird für die radiologische Auswertung ein Sicherheitsabstand von ≥ 10 mm für die invasive Komponente und ≥ 2 mm für die nicht invasive Komponente bezüglich der bildgebenden Diagnose R0/R1 festgelegt. Ein größerer als in der Histologie geforderter Abstand beim invasiven Karzinom wurde gewählt, um ein möglicherweise assoziiertes peritumorales, nicht-kalktragendes DCIS zu inkludieren bei gleichzeitig in der Regel gut abgrenzbaren MK im Falle eines kalktragenden DCIS.

Um einen validen, direkten Vergleich der R0/R1-Diagnose aus den radiologischen Verfahren und aus der Histologie zu ermöglichen wird angestrebt, möglichst gleiche Randbedingungen zur Ableitung der Diagnosen zu verwenden. Deshalb wird nicht die originale endgültige R0/R1-Diagnose der Histologie für den Vergleich verwendet, sondern auf Basis der histologisch am Präparat, ohne

mögliche intraoperativ gewonnene Nachresektate, gemessenen Randabstände eine neue Diagnose abgeleitet. Da die Nachresektate radiographisch nicht dokumentiert werden, ist der Vergleich andernfalls nicht valide. Darüber hinaus werden bei der Ableitung der neuen histologischen R0/R1-Diagnose auch dieselben Randabstandskriterien (≥ 10 mm bzw. ≥ 2 mm) wie in der Auswertung der radiologischen Verfahren genutzt. Nur so kann ein valider direkter Vergleich zwischen den Diagnosen der Methoden sichergestellt werden.

2.4 Bewertungsbögen

In der ersten Auswertungsphase ist die präoperative Mammographie und die intraoperative 2D-Präparateradiographie von zwei Readern unabhängig voneinander und bezogen auf die Daten der Patientinnen und Histologie verblindet, analysiert worden. Bekannt war die Malignität. Pro Fall ist die präoperative Mammographie anhand des Bewertungsbogens 1 (siehe Anhang 8.1) und in gleicher Sitzung die PR anhand des Bewertungsbogens 2 (siehe Anhang 8.2) beurteilt und vermessen worden. Nach einer „Auswaschphase“ von mindestens 30 Tagen wurde das Bildmaterial der DBT mit Hilfe des Bewertungsbogens 3 (siehe Anhang 8.3) bearbeitet. Die präoperativen Mammographiebilder sowie die entsprechenden Bewertungsbögen durften während der Auswertung der PR und der DBT begutachtet werden. Es gab keine zeitliche Beschränkung für die Beurteilung der verschiedenen Verfahren. Die Reader stoppten stichprobenhaft die Bearbeitungsdauer der jeweiligen Auswertungsphasen, um einen Vergleich zum subjektiv abgefragten zeitlichen Aufwand erstellen zu können.

2.5 Verwendete mathematische Methoden

2.5.1 Konkordanzanalyse

Zur Beurteilung der Übereinstimmung von stetigen Messgrößen werden sowohl für den Interreader- als auch für den Intermethodenvergleich Korrelations- / Streudiagramme sowie Bland-Altman- Diagramme verwendet.

In den Korrelations- / Streudiagrammen sind die Messwerte zweier Reader oder zweier Methoden gegeneinander aufgetragen. Zur visuellen Kontrolle der

Assoziation zwischen den Messergebnissen ist das Ergebnis der Regressionsanalyse nach Passing-Bablok [84] in Form einer Regressionsgerade dargestellt. Zur Visualisierung der Streuung sind ebenfalls die Regressionsgeraden dargestellt, die sich aus den 95%-Konfidenzintervallen ergeben. Zur Berechnung von Konfidenzintervallen wurde grundsätzlich ein Signifikanzniveau von 5 % festgelegt.

Die Bland-Altman-Diagramme [85, 86] zeigen die absoluten Differenzen der Wertepaare über dem Mittelwert beider Messungen. Zusätzlich sind die mittlere Abweichung aller Messwerte sowie die Limits of Agreement (LoA) eingezeichnet. Die mittlere Abweichung kann als systematische Abweichung interpretiert werden, die gegebenenfalls korrigierbar ist. Die LoA werden aus dem Mittelwert $\pm 1,96$ -Standardabweichungen der Stichprobe gebildet, da bei Annahme eines normalverteilten Fehlers 95 % aller Differenzen der Grundgesamtheit innerhalb von $\pm 1,96$ -Standardabweichungen der Stichprobe erwartet werden. Damit kennzeichnen die LoA den Übereinstimmungsbereich der Messungen. Sie sind ein aus der Stichprobe abgeleiteter Schätzer für die erwartete (zufällige) Streuung in der Grundgesamtheit. Diese Streuung ist in der Regel nicht korrigierbar. Enge LoA sind somit im Sinne eines Interreadervergleichs Grundvoraussetzung für eine konsistente Messmethodik und geben auch bei Intermethodenvergleichen eine Indikation für eine hohe Übereinstimmung. Die maximal zulässige Größe der LoA kann jedoch nicht mathematisch bestimmt werden, sie ist vielmehr abhängig vom klinischen Kontext [85, 86].

Im Hinblick auf die klinische Relevanz wird im Rahmen dieser Arbeit ein maximal zulässiger Übereinstimmungsbereich von ± 5 mm um den Mittelwert der Differenzen herum festgelegt, da bei oben beschriebenen Forderungen an die radiologischen Randabstände eine höhere Abweichung zu einer unzuverlässigen Aussage bezüglich der erfolgreichen Tumorsektion führen würde.

Die Beurteilung der Übereinstimmung von Messwerten auf nominalen Skalen (insb. R0/R1 im Intermethoden- und Interreader-Vergleich) wurden anhand von Kontingenztafeln sowie anhand der zufallskorrigierten Übereinstimmungsmaßzahl Cohens Kappa (K_2) durchgeführt. Cohens Kappa repräsentiert dabei den normierten Anteil der Übereinstimmung, der über die zufallsbasierte

Übereinstimmung hinausgeht [87]. Die Maßzahl K2 kann negative Werte und Werte bis 1 annehmen. Dabei entspricht der Wert $K2 = 1$ einer perfekten Übereinstimmung. Bei einem Wert von 0 liegt die Übereinstimmung auf dem Niveau einer zufälligen Übereinstimmung bei gegebenen Randhäufigkeiten der Stichprobe. Negative Werte weisen eine schlechtere Übereinstimmung als eine rein zufällige Übereinstimmung auf. Gemäß [86] können folgende Interpretationen aus den Wertebereichen für K2 aus Tabelle 1.1 abgeleitet werden:

Tabelle 2.1: Beurteilung der Übereinstimmung anhand von Cohens Kappa

Wert von K2	Beurteilung der Übereinstimmung
$\leq 0,20$	nicht ausreichend
0,21-0,40	hinreichend
0,41-0,60	moderat
0,61-0,80	gut
0,81-1,0	sehr gut

Beim Vergleich von radiologischen Ergebnissen mit dem Goldstandard, der Histologie, wird aus den Kontingenztafeln zusätzlich die Sensitivität sowie die Spezifität berechnet. Die Sensitivität gibt dabei den Anteil der durch das radiologische Verfahren aus Sicht der Histologie richtig identifizierten R1-Befunde an. Die Spezifität beziffert hingegen den Anteil der durch das radiologische Verfahren aus Sicht der Histologie richtig identifizierten R0-Befunde.

2.5.2 Verwendete Software

Die Daten der bearbeiteten Auswertungsbögen wurden in eine Excel-Tabelle übertragen. Es wurde pro Verfahren für jeden Reader ein Tabellenblatt erstellt, sowie eines für die Histologie. Zur weiteren statistischen Verarbeitung wurden für jedes Auswertungskriterium eindeutige Variablen definiert. Die statistische Auswertung erfolgte mit der Software MATLAB (MathWorks Inc., R2019b).

3 Ergebnisse

3.1 Studienpopulation und -kollektiv

In diesem Abschnitt wird die Stichprobe der Präparate aus dieser Arbeit anhand der Merkmale der Altersstruktur und der Entität beschrieben, sowie mit größeren Datensätzen aus der Literatur verglichen.

Abbildung 3.1 zeigt die Verteilung des Alters der Patientinnen aus Datensatz 1 zum Zeitpunkt der Operation. Die jüngste Patientin ist 40 Jahre alt, die älteste 83 Jahre. Im Mittel sind die Patientinnen 60,2 Jahre alt, der Median liegt bei 59,5 Jahren.

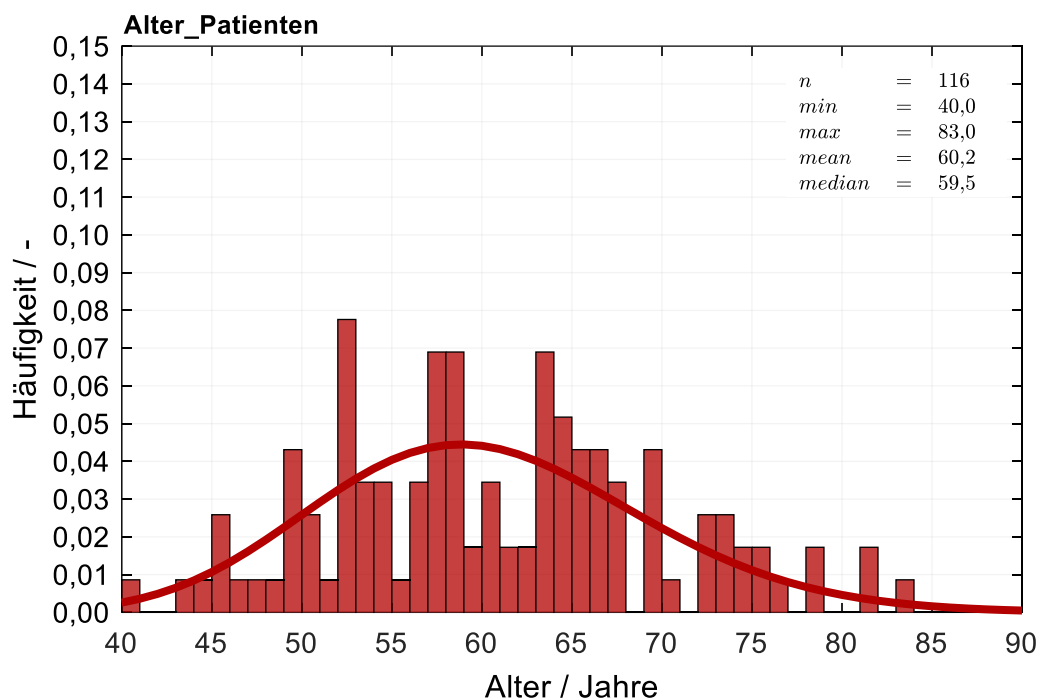


Abbildung 3.1: Histogramm Altersverteilung (Datensatz 1)

Zum Vergleich liegt das mediane Erkrankungsalter von Frauen in Baden-Württemberg bei 64 Jahren [88]. Tabelle 3.1 vergleicht die Altersverteilung in vorliegender Studie (Datensatz 1, $n = 124$) mit denen der Grundgesamtheit von Patientinnen aus Baden-Württemberg in den Jahren 2014 & 2015 ($n = 17.062$). Gegenüber der Grundgesamtheit des Erkrankungsalters in Baden-Württemberg ist die Altersgruppe ≤ 64 Jahre in dieser Studie eher überrepräsentiert, die Altersgruppe ≥ 80 hingegen unterrepräsentiert.

Tabelle 3.1: Vergleich Anteile nach Altersgruppen mit Daten aus dem Krebsregister Baden-Württemberg

Altersgruppen	Baden-Württemberg 2014 & 2015	Vorliegende Studie
≤64	52,9 %	63,8 %
65-79	34,4 %	33,6 %
≥80	12,7 %	2,6 %

Abbildung 3.2 zeigt die Einteilung nach Entitäten aus den histologischen Befunden. Zum größten Teil handelt es sich um duktales Karzinome (NST). In der vorliegenden Studie ist zusätzliches DCIS häufig mit der invasiven Komponente assoziiert (60,3 %).

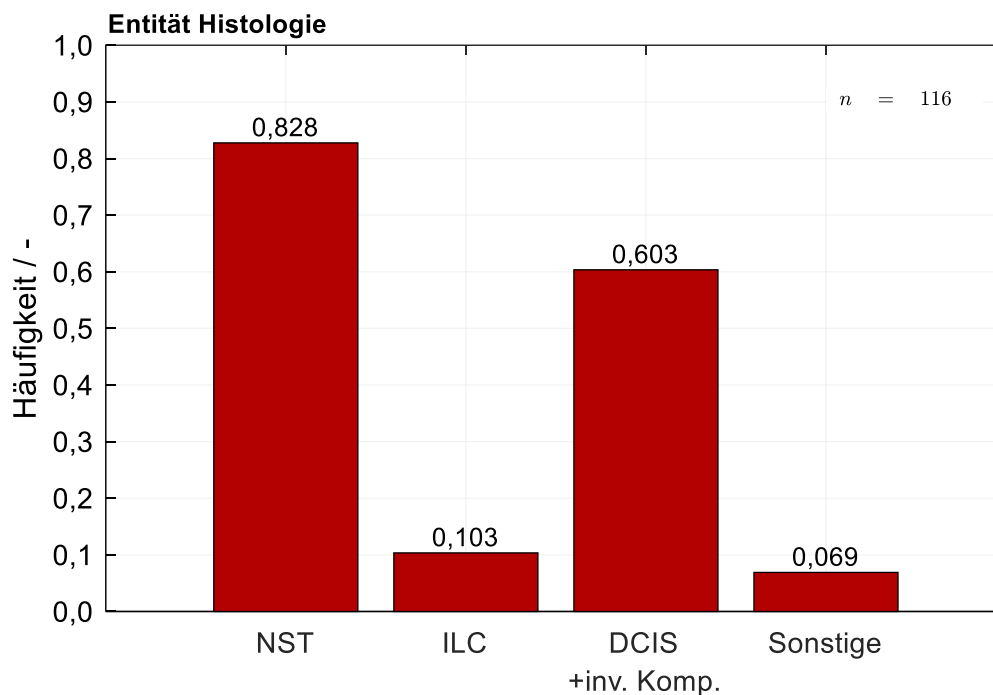


Abbildung 3.2: Anteile nach Entität in der Stichprobe

Ein Vergleich mit Literaturdaten [5, 11] zeigt eine gute Übereinstimmung bezüglich der Verteilung der Entitäten aus der Stichprobe von Datensatz 1 (siehe Tabelle 3.2). Darüber hinaus wird in der Literatur ebenfalls beschrieben, dass DCIS in bis zu 85 % mit dem Befund eines invasiven Karzinoms assoziiert ist [22]. In Summe zeigt sich bezüglich der betrachteten Merkmale eine gute Übereinstimmung zu anderen Literaturdaten, sodass nicht von einer systematischen Verzerrung auszugehen ist.

Tabelle 3.2: Vergleich der Verteilung der Entität mit Literaturdaten

	Böcker	Petru	Vorliegende Studie
NST	ca. 60 %	65-80 %	82,8 %
ILC	10-20 %	ca. 10 %	10,3 %
Sonstige	3-10 %	/	6,9 %
DCIS + inv. Komp.	/	/	60,3 %

3.2 Ergebnisse präoperative Mammographie und Inter-Reader-Vergleich

Der nachfolgende Abschnitt beschreibt die wesentlichen Ergebnisse aus der Auswertung der Bewertungsbögen zur präoperativen Mammographie. Neben der Verteilung der Brustdichte, der Lokalisation der Herdbefunde, wird insbesondere auf die Ergebnisse des Inter-Reader-Vergleichs und auf die Entität eingegangen.

Brustdichte, Lokalisation und Entität

Die Verteilung der Brustdichte nach ACR, die im Rahmen der Befundung der Mammographie ermittelt wurde, ist in Abbildung 3.3 dargestellt.

In der vorliegenden Arbeit ist die Dichte der Brust zu ca. 80...90 % in Kategorie ACR B & ACR C verortet. Vergleichswerte aus der Literatur [89] zeigen eine ähnliche Verteilung auch bei deutlich größeren Fallzahlen (n: 111456, Alter: 50-69). Tabelle 3.3 stellt diese mit den Daten aus vorliegender Studie gegenüber, wobei hier die Spannweite beider Reader angegeben wurde. Die meisten Mammakarzinome der Stichprobe sind im oberen äußeren Quadranten der Brust lokalisiert (links: 35,3-36,2 %, rechts: 28,4-31,9 %). Diese Aussage deckt sich mit Angaben in der Literatur [5, 11].

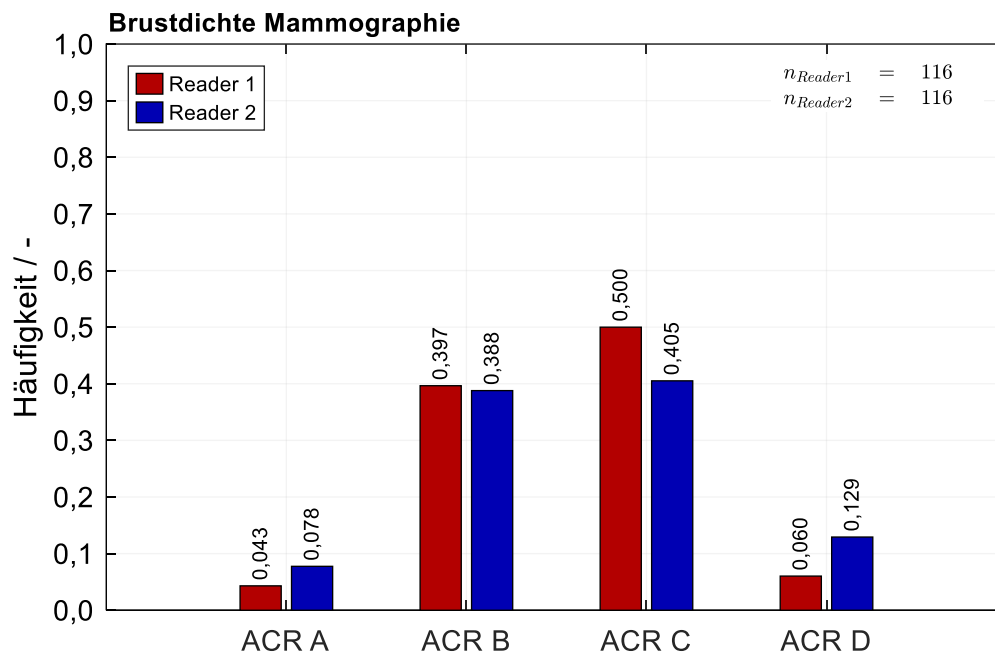


Abbildung 3.3: Anteil der Brustdichte in der Stichprobe (Datensatz 1)

Tabelle 3.3: Vergleich der Verteilung der Brustdichte mit Literaturdaten

	Weigel	Vorliegende Studie
ACR A	6 %	4,3-7,8 %
ACR B	44 %	38,8-39,7 %
ACR C	45 %	40,5-50,0 %
ACR D	5 %	6,0-12,9 %

Abbildung 3.4 zeigt die Einteilung der Entitäten beider Reader verglichen mit dem Ergebnis der Histologie. Der hohe Anteil an DCIS aus der Histologie findet sich in der Bewertung der Reader auf Basis der Mammographie nicht wieder. Dies könnte zum einem an fehlenden MK im DCIS liegen oder daran, dass das DCIS im Bereich des invasiven Herdbefundes lokalisiert und somit schwer von diesem abgrenzbar ist, da auch ein rein invasives Karzinom mit suspekten MK einhergehen kann.

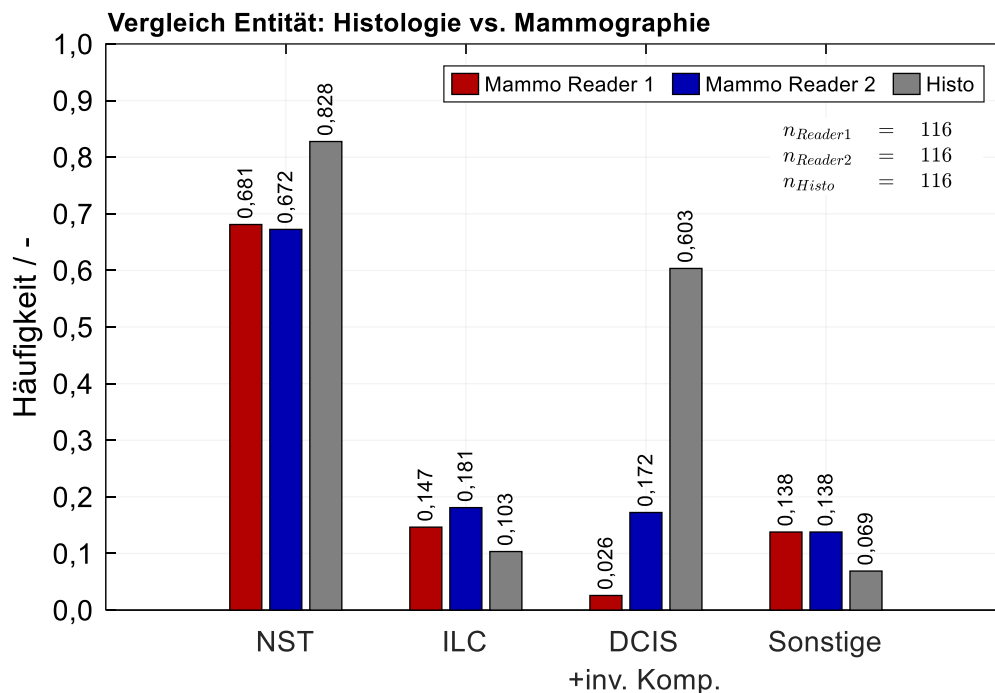


Abbildung 3.4: Vergleich der Entität aus Mammographie mit der Histologie (Datensatz 1)

Inter-Reader-Vergleich Herdbefundgröße

Zur Bewertung der Inter-Reader-Reliabilität der Methode Mammographie bei der Bestimmung der Herdbefundgröße sowie für die Messung der Größe von MK wurde die in Abschnitt 2.5.1 beschriebene Methode nach Bland-Altman kombiniert mit der Regressionsanalyse nach Passing-Bablok verwendet. Abbildung 3.5 zeigt beispielhaft die Übereinstimmung der Messungen beider Reader bei der Größe CCx. Auf Basis des Korrelationsdiagramms (links) ist eine grundsätzliche Korrelation zwischen den Messungen der Reader zu erkennen. Ausreißer liegen zu beiden Seiten der Winkelhalbierenden (perfekte Übereinstimmung) vor. Gemäß der Methode nach Passing Bablok kann keine vollständige Übereinstimmung der Reader festgestellt werden, da zwar die Steigung der Regressionsgerade (β) den Wert 1 aufweist sowie das 95%-Konfidenzintervall ($\beta_{KI95\%}$) den Wert 1 umschließt, jedoch der y-Achsenabschnitt der Regressionsgerade (α) sowie das zugehörige 95%-Konfidenzintervall ($\alpha_{KI95\%}$) größer null

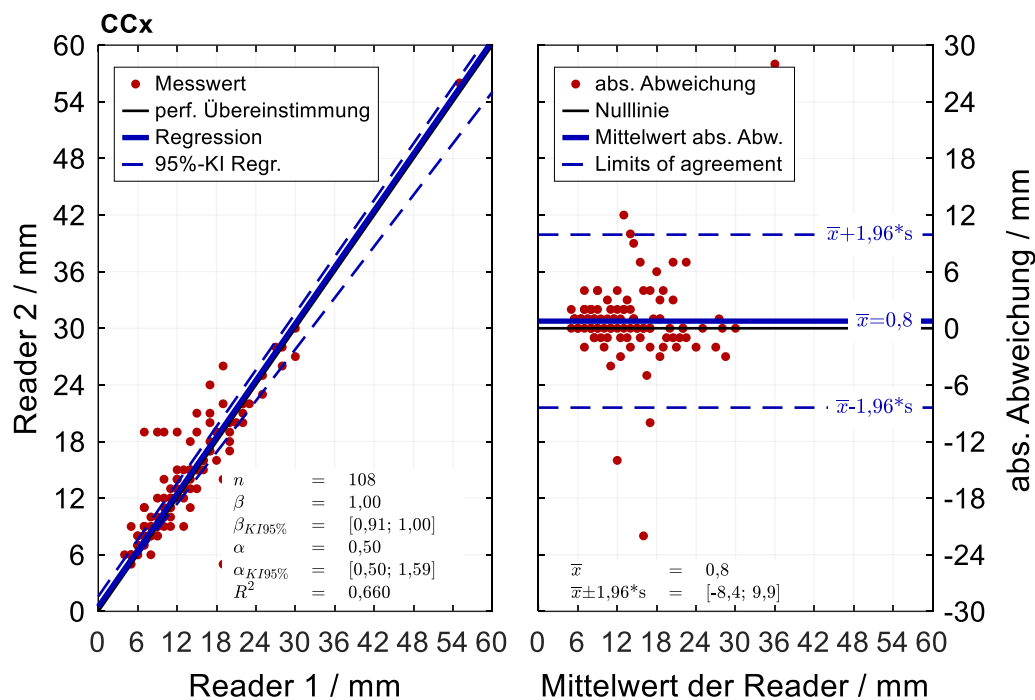


Abbildung 3.5: Inter-Reader-Vergleich der Größe des Herdbefundes am Beispiel CCx (MG)

sind. Damit liegt ein – wenn auch kleiner – Versatz in der Regressionsanalyse vor. Um zusätzlich zur Regression auch die tatsächliche Güte der Übereinstimmung darzustellen, sind im rechten Diagramm die absoluten Abweichungen zwischen den Messwerten über deren Mittelwert abgebildet (Bland-Altman-Diagramm, rechts). Die mittlere Abweichung zwischen den Messungen beider Reader liegt bei 0,8 mm. Die Limits-of-Agreement (LoA), also die Intervallgrenzen innerhalb derer bei Annahme standardnormalverteilter Messdifferenzen 95 % der Messstreuung der Grundgesamtheit zu erwarten sind, liegen bei -8,4 bis 9,9 mm. Im Vergleich zum Messbereich von ca. 0-60 mm erscheint diese 95%-Streubreite mit 18,3 mm hoch. Die LoA liegen zudem außerhalb der mit Blick auf die klinische Relevanz im Rahmen dieser Arbeit festgelegten maximal zulässigen LoA von ± 5 mm. Es kann daher für die Messgröße CCx zwar eine im Mittel gute Übereinstimmung bei jedoch hoher Streuung festgehalten werden. Eine vollständige Übereinstimmung oder Austauschbarkeit der Reader kann nach Passing-Bablok für diesen Fall nicht abgeleitet werden. Tabelle 3.4 zeigt die Vergleichsgrößen für die weiteren erhobenen Messgrößen.

Tabelle 3.4: Inter-Reader-Vergleichswerte Größe des Herdbefundes (MG)

	Mittlere Abwei- chung	Standardabw. Differenzen	LoA	
	mm	mm	mm	
CCx	0,7593	4,6762	-8,4061	9,9246
CCz	1,3173	4,5181	-7,5381	10,1727
MLOy	0,2075	5,0061	-9,6045	10,0195
MLOz	1,2547	4,3928	-7,3551	9,8645

Die jeweiligen Diagramme können im Anhang 8.4 eingesehen werden. Die Analyse der Übereinstimmung auf Basis der weiteren Messgrößen des Herdbefundes bestätigt die oben beschriebenen Erkenntnisse des Inter-Reader-Vergleichs.

Vorkommen und Inter-Reader-Vergleich Mikrokalzifikationen

Abbildung 3.6 vergleicht den Anteil von Patientinnen, bei denen von den Readern MK identifiziert wurde, sowie dessen Lokalisation innerhalb bzw. außerhalb des Herdbefundes. Es ist ersichtlich, dass in der Mammographie bei 21,6-25,0 % der Fälle durch die Reader MK identifiziert wurde. Der größere Anteil findet sich innerhalb des Herdes. Der Vergleich mit den Ergebnissen der Histologie zeigt, dass der Anteil an Fällen bei denen MK vorliegt, geringer ist (13,8 %, 16 von 116 Fällen in Datensatz 1).

Der Inter-Reader-Vergleich auf Basis der Größe der Mikrokalzifikationen für die Messebene MLOy ist in Abbildung 3.7 dargestellt. Auf Basis des Korrelationsdiagramms (links) ist zwar grundsätzlich eine Assoziation zwischen den Messungen der Reader zu erkennen, jedoch liegen die Ausreißer nicht symmetrisch zur Winkelhalbierenden, sodass ein y-Achsenabschnitt der Regressionsgerade größer 1 resultiert. Aufgrund der geringen Anzahl der Messgrößen weisen die Konfidenzintervalle große Werte auf. Eine Übereinstimmung für die Messebene MLOy ist daher nur schwer zu beurteilen.

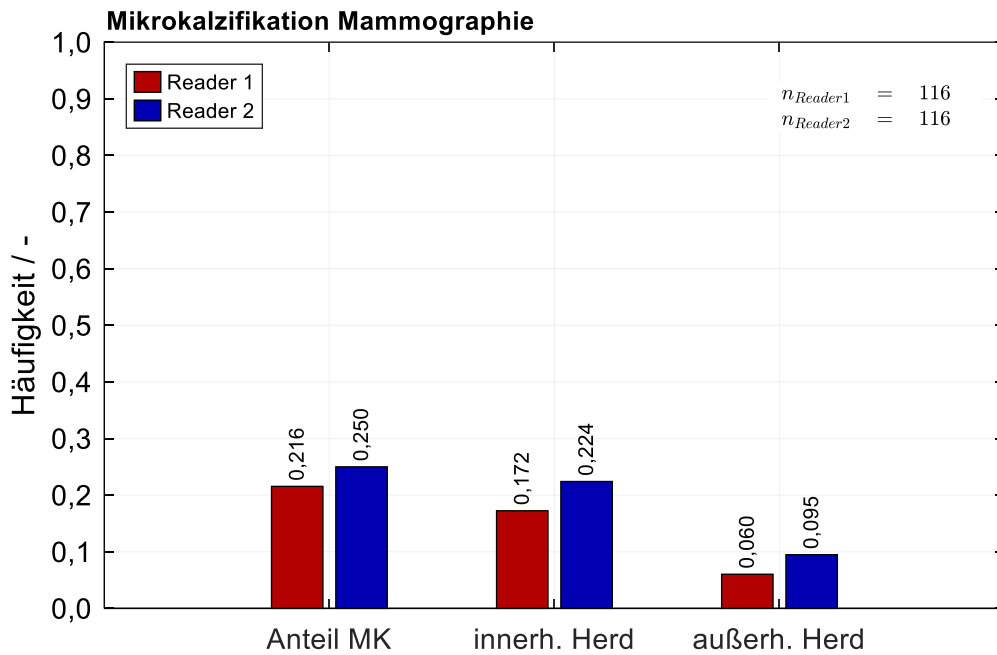


Abbildung 3.6: Anteil MK in der Stichprobe (MG)

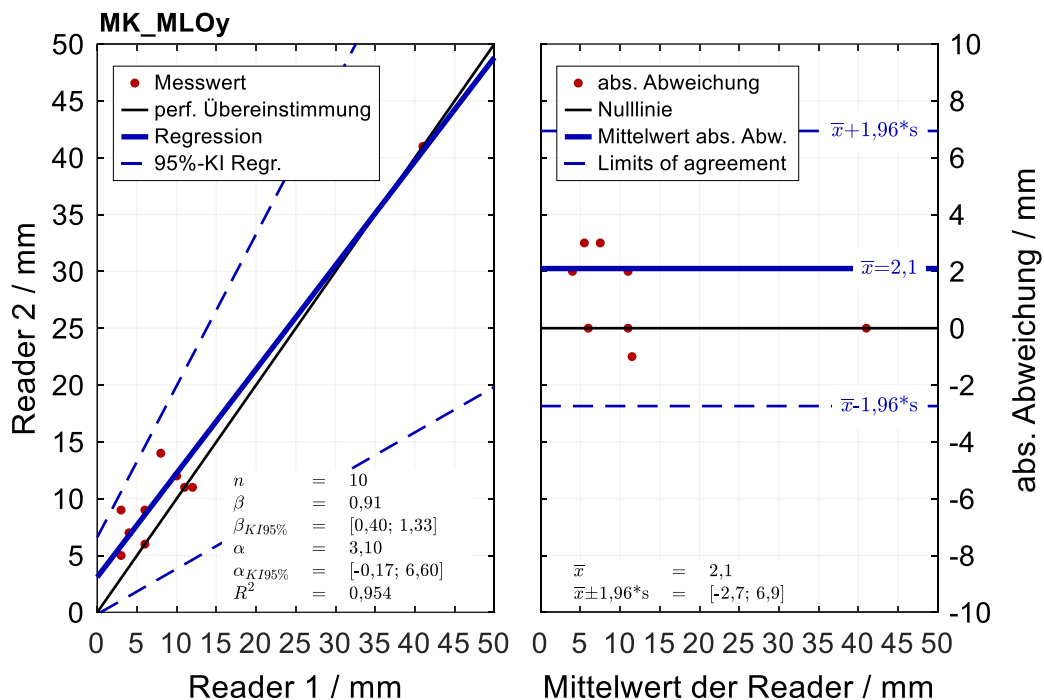


Abbildung 3.7: Inter-Reader-Vergleich der Größe der MK am Beispiel MLOy (MG)

Die mittlere Abweichung zwischen den Messungen beider Reader liegt bei 2,1 mm (Bland-Altman-Diagramm, rechts). Die LoA liegen bei -2,7 bis 6,9 mm. Die im Rahmen dieser Arbeit angesetzten maximal zulässigen LoA werden verletzt.

Tabelle 3.5 zeigt die Vergleichswerte für die weiteren erhobenen Messgrößen. Die jeweiligen Diagramme können im Anhang 8.4 eingesehen werden.

Tabelle 3.5: Inter-Reader-Vergleichswerte Größe MK (MG)

	Mittlere Abweichung	Standardabw. Differenzen	LoA	
	mm	mm	mm	
MK_CCx	2,3750	2,8253	-3,1625	7,9125
MK_CCz	13,5714	20,9512	-27,492	54,635
MK_MLOy	2,1000	2,4698	-2,7408	6,9408
MK_MLOz	7,2000	15,9360	-24,0345	38,4345

Die Analyse der Übereinstimmung auf Basis der weiteren Messgrößen des Mikrokalks bestätigt die oben beschriebenen Erkenntnisse des Inter-Reader-Vergleichs. Die MK-Messgrößen variieren zwischen den Readern stark, die mittlere Abweichung ist im Maximum groß (13,6 mm), gleiches gilt für die LoA, die alle oberhalb der im Rahmen dieser Arbeit festgelegten maximal zulässigen LoA liegen. Eine abschließende Bewertung der Inter-Reader-Reliabilität erfordert jedoch einen größeren Stichprobenumfang, der im Rahmen dieser Arbeit nicht abgebildet werden kann.

3.3 Ergebnisse intraoperative PR und Inter-Reader-Vergleich

Der nachfolgende Abschnitt stellt die relevanten Ergebnisse aus der Auswertung der Bewertungsbögen zur intraoperativen PR dar. Hierbei wird insbesondere auf die Ergebnisse des Inter-Reader-Vergleichs der Messwerte zur Präparatgröße, zur Größe des Herdbefundes sowie zum Randabstand eingegangen. Die durch die Reader abgeleiteten Diagnosen (R0/R1) werden mittels Kontingenztafeln und Cohens Kappa verglichen.

3.3.1 Präparatgröße

Im Vergleich zu den Messgrößen von Herdbefund und Randabstand wird für die Präparatgröße die kleinste Streuung zwischen den Readern erwartet. Diese Annahme basiert auf der zum einen trennschärfsten Begrenzung des Bildmaterials und zum anderen auf der Größe der zu messenden Werte. Abbildung 3.8 zeigt den Inter-Reader-Vergleich der Messungen beider Reader zur Präparatgröße am Beispiel APx.

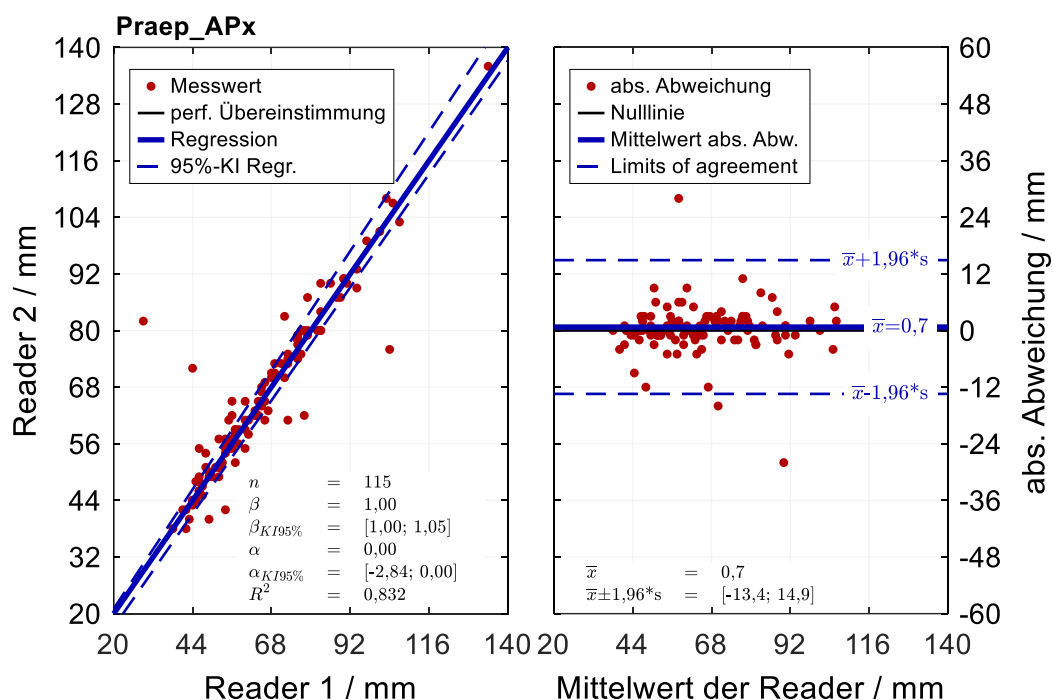


Abbildung 3.8: Inter-Reader-Vergleich der Größe des Präparats am Beispiel APx (PR)

Auf Basis des Korrelationsdiagramms (links) ist eine grundsätzlich gute Korrelation zwischen den Messungen der Reader zu erkennen. Ausreißer liegen zu beiden Seiten der Winkelhalbierenden vor. Gemäß der Methode nach Passing Bablok besteht formal eine vollständige Übereinstimmung der Reader, da β den Wert 1 und α den Wert 0 aufweist, sowie die zugehörigen 95%-Konfidenzintervalle die Werte 1 bzw. 0 umschließen. Die mittlere Abweichung zwischen den Messungen beider Reader liegt bei 0,7 mm. Die LoA liegen bei -13,4...14,9 mm. Im Vergleich zum Messbereich von ca. 0-140 mm erscheint diese 95%-Streuweite mit 28,3 mm hoch. Es kann daher eine gute Korrelation mit formal

vollständiger Übereinstimmung nach Passing Bablok bei jedoch hoher Streuung festgehalten werden.

Tabelle 3.6 zeigt die Vergleichsgrößen für die weiteren erhobenen Messgrößen zur Präparatgröße. Die jeweiligen Diagramme können im Anhang 8.5 eingesehen werden.

Tabelle 3.6: Inter-Reader-Vergleichswerte Größe des Präparats (PR)

	Mittlere Abweichung	Standardabw. Differenzen	LoA	
	mm	mm	mm	
Praep_APx	0,7391	7,2306	-13,4329	14,9112
Praep_APz	1,6087	7,7241	-13,5305	16,7479
Praep_LLy	0,1754	5,7287	-11,0529	11,4038
Praep_LLz	1,3860	7,4662	-13,2477	16,0196

Die Analyse der Übereinstimmung auf Basis der weiteren Messgrößen der Präparatgröße bestätigt die oben beschriebenen Erkenntnisse des Inter-Reader-Vergleichs.

3.3.2 Größe Herdbefund

Abbildung 3.9 stellt den Vergleich der Messungen beider Reader bezüglich der Größe des Herdbefundes am Beispiel APx dar. Auf Basis des Korrelationsdiagramms ist eine Korrelation zwischen den Messungen der Reader zu erkennen. Ausreißer liegen tendenziell bei größeren Werten von Reader 2, was sich in einem positiven y-Achsenabschnitt der Regressionsgeraden niederschlägt. Entsprechend kann nach Passing Bablok keine vollständige Übereinstimmung der Reader festgestellt werden. Zwar weist die Steigung der Regressionsgerade den Wert 1 auf und das 95%-Konfidenzintervall umschließt den Wert 1, jedoch sind der y-Achsenabschnitt der Regressionsgerade sowie das zugehörige 95%-Konfidenzintervall größer null. Die mittlere Abweichung zwischen den Messungen beider Reader liegt bei 1,4 mm (Abbildung 3.9, rechts). Die LoA liegen bei -8,3...11,1 mm. Im Vergleich zur Messung der Präparatgröße ist die absolute 95%-Streubreite im vorliegenden Fall mit 19,4 mm zwar kleiner, mit Blick auf den

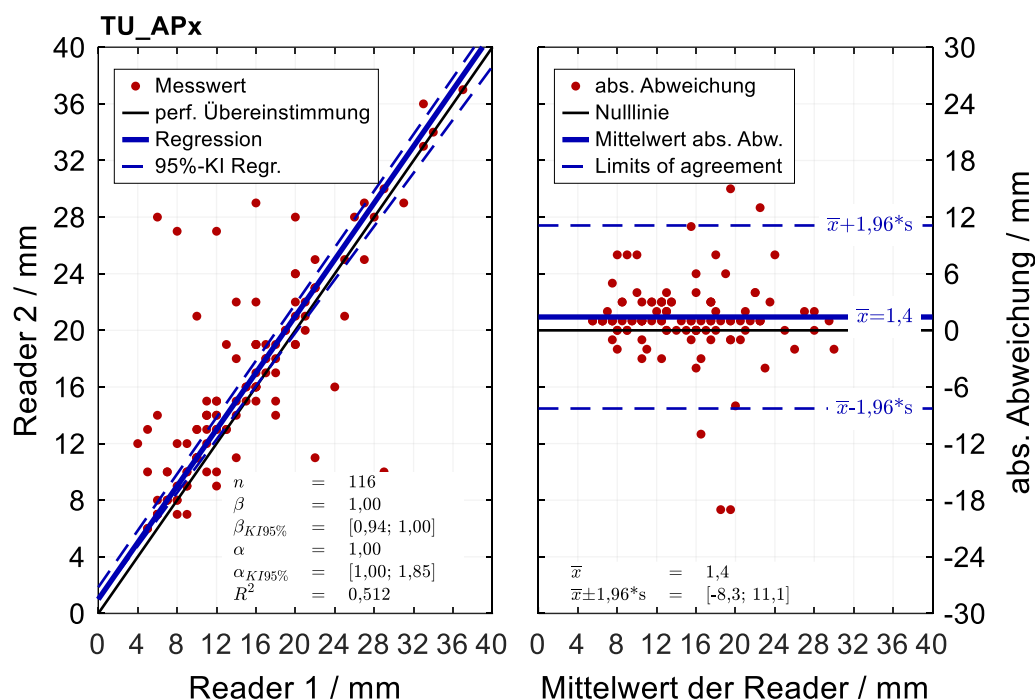


Abbildung 3.9: Inter-Reader-Vergleich der Größe des Herdbefundes am Beispiel APx (PR)

kleineren Messbereich von ca. 0-40 mm jedoch relativ gesehen höher und somit insgesamt als hoch einzustufen.

Tabelle 3.7 zeigt die Vergleichsgrößen für die weiteren erhobenen Messungen der Herdbefundgröße. Die jeweiligen Diagramme können im Anhang 8.5 eingesehen werden.

Tabelle 3.7: Inter-Reader-Vergleichswerte Größe des Herdbefundes (PR)

	Mittlere Abweichung	Standardabw. Differenzen	LoA	
	mm	mm	mm	
TU_APx	1,4138	4,9486	-8,2854	11,1130
TU_APz	1,3103	6,7649	-11,9488	14,5695
TU_LLy	1,0783	5,6338	-9,9639	12,1205
TU_LLz	0,8957	6,1621	-11,1821	12,9734

Die Analyse der Übereinstimmung auf Basis der weiteren Messgrößen des Herdbefundes bestätigt die oben beschriebenen Erkenntnisse des Inter-Reader-Vergleichs. Erwartungsgemäß ergibt sich im Vergleich zur Messung der

Präparatgröße eine relativ gesehen größere Streuung zwischen den Readern, da der Herdbefund eine geringere Ausdehnung besitzt und die optische Trennschärfe bzw. Abgrenzbarkeit geringer ist.

3.3.3 *Randabstand*

In der Theorie können die Randabstände in 10 Betrachtungsebenen gemessen werden (vgl. Bewertungsbogen in Abschnitt 8.2 und 8.3), wobei pro Präparat immer nur 8 Abstände gemessen werden können (je nach Lage des Präparats können medialer und lateraler oder kranialer und kaudaler Abstand gemessen werden). Je nach Lokalisation des Präparats und der Randabstände wählen die Reader die zutreffenden Messgrößen aus. Mitunter ist in vorliegender Arbeit zu beobachten, dass die Reader nicht alle Randabstandsvariablen gleich wählen. Da nur die Messwerte verglichen werden können, zu denen die Messwerte beider Reader vorliegen, ergeben sich zum Teil Fallzahlen, die unter dem Stichprobenumfang von Datensatz 1 liegen.

Abbildung 3.10 zeigt den Vergleich der Übereinstimmung der Messungen beider Reader für die Messgröße des Randabstandes am Beispiel APkaudal, wobei der Stichprobenumfang hier $n = 61$ beträgt. Im Korrelationsdiagramm ist eine grundsätzliche Korrelation zwischen den Messungen der Reader zu erkennen, wobei Ausreißer tendenziell bei größeren Werten von Reader 2 liegen. Der Versatz macht sich entsprechend in einem positiven y-Achsenabschnitt der Regressionsgerade bemerkbar. Somit ist trotz einer Steigung von 1 keine vollständige Übereinstimmung feststellbar. Die mittlere Abweichung zwischen den Messungen beider Reader liegt bei 0,9 mm und kann als klein eingeschätzt werden. Die LoA liegen bei -6,0...7,8 mm. Eine 95%-Streubreite von 13,8 mm ist im direkten Vergleich zur Größe des Herdbefund und des Präparats klein und erscheint für den Messbereich von ca. 0-70 mm als moderat. Trotzdem wird das in dieser Arbeit festgelegte klinisch relevante Gütekriterium von ± 5 mm überschritten. Es kann daher zwar eine grundsätzliche Korrelation mit moderater Streubreite, aber keine vollständige Übereinstimmung oder Austauschbarkeit der Reader abgeleitet werden. Die Vergleichsgrößen für die weiteren Messgrößen des Randabstandes in den anderen Ebenen sind in Tabelle 3.8 dargestellt.

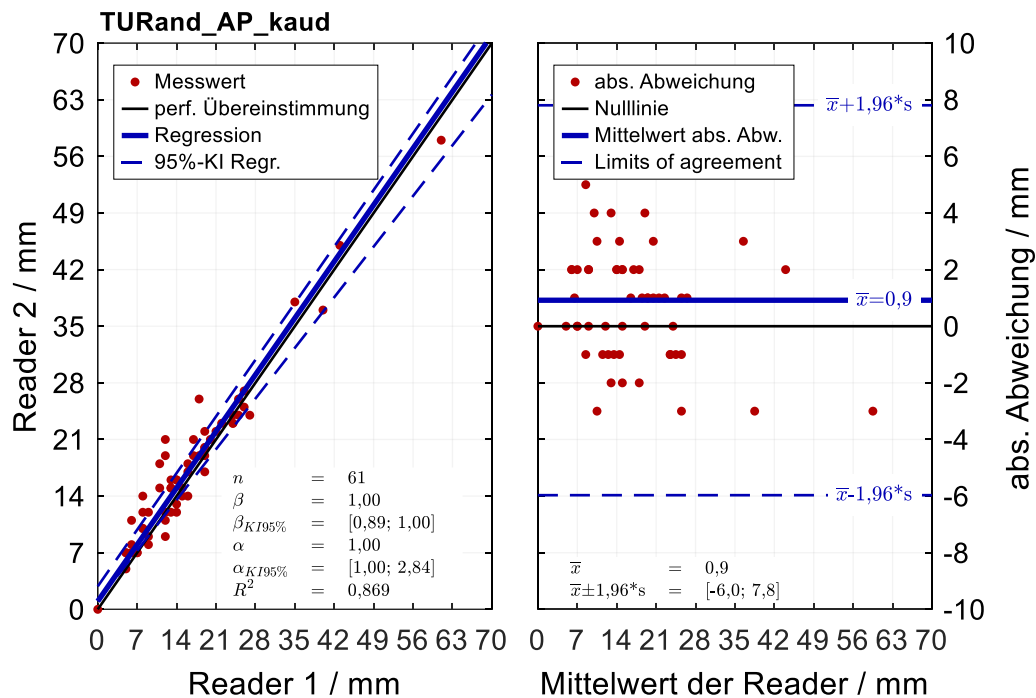


Abbildung 3.10: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel APkaudal (PR)

Tabelle 3.8: Inter-Reader-Vergleichswerte Größe des Randabstandes (PR)

	n	Mittlere Abweichung	Standardabw. Differenzen	LoA	
	-	mm	mm	mm	
AP_mam	101	2,3168	7,4912	-12,3660	16,9997
AP_per	105	-0,9238	7,3651	-15,3593	13,5117
AP_kran	63	0	6,4006	-12,5452	12,5452
AP_kaud	61	0,9180	3,5133	-5,9680	7,8041
AP_med	26	-1,3077	4,7562	-10,6299	8,0145
AP_lat	27	-1,4444	5,7935	-12,7996	9,9107
LL_dors	102	-0,1765	4,7375	-9,4619	9,1090
LL_kut	108	-0,5093	4,6554	-9,6339	8,6153
LL_mam	103	0,5243	11,1460	-21,3218	22,3703
LL_per	105	-2,8286	10,7117	-23,8234	18,1663

Die jeweiligen Diagramme können im Anhang 8.5 eingesehen werden. Die Analyse der Übereinstimmung auf Basis der weiteren Messgrößen des

Randabstandes bestätigt im Grunde die oben beschriebenen Erkenntnisse des Inter-Reader-Vergleichs, wobei die LoA in den anderen Ebenen mitunter höher ausfallen. Die mittlere Abweichung ist über alle Messgrößen verhältnismäßig klein, wenngleich auch hier eine moderate bis hohe Streuung vorliegt.

3.3.4 Inter-Reader-Vergleich der Diagnose R0/R1

In Tabelle 3.9 ist der Vergleich der abgeleiteten Diagnosen der beiden Reader in Form zweier Kontingenztafeln, einmal absolut und einmal relativ, dargestellt.

Tabelle 3.9: Kontingenztafel R0/R1-Vergleich intraoperative PR

Kontingenztafel intraoperative PR R0/R1 (abs.)

		Reader 1		Randhäufigkeit
		R0	R1	
Reader 2	R0	47	14	61
	R1	11	44	55
Randhäufigkeit		58	58	116

Kontingenztafel intraoperative PR R0/R1 (rel.)

		Reader 1		Randhäufigkeit
		R0	R1	
Reader 2	R0	0,40517	0,12069	0,52586
	R1	0,09483	0,37931	0,47414
Randhäufigkeit		0,5	0,5	

p0 0,78448

pe 0,5

K2 0,56897

unterer Wert 0,41931 **oberer Wert** 0,71862

Die Kontingenztafel zeigt die jeweilige Anzahl bzw. Häufigkeit der Kombinatorik aus den möglichen Urteilsausprägungen der beiden Reader. Zusätzlich sind die Randhäufigkeiten, der Stichprobenumfang sowie die in Abschnitt 2.5.1 beschriebene Übereinstimmungsmaßzahl Cohens Kappa (K2) abgebildet. Darüber hinaus finden sich zusätzlich die zur Berechnung von K2 notwendigen Größen der geschätzten Wahrscheinlichkeit für eine übereinstimmende Beurteilung

beider Reader (p_0) sowie der zu erwartenden Wahrscheinlichkeit einer Urteilsübereinstimmung bei gegebenen Randhäufigkeiten und bei stochastisch unabhängigen Beurteilern (p_e). Zusätzlich ist das 95%-Konfidenzintervall (oberer/unterer Wert) für K_2 angegeben.

Es besteht eine übereinstimmende Beurteilung beider Reader für R0/R1 in 91 von 116 Fällen aus Datensatz 1 (78 %). Die Inter-Reader-Reliabilität ($K_2 = 0,569$) muss gemäß Altman [86] als moderat eingestuft werden. Das 95%-Konfidenzintervall liegt im moderaten bis guten Bereich. Die moderate Übereinstimmung kann dabei durch die Streuung der Messwerte des Randabstandes zwischen den Readern erklärt werden.

3.4 Ergebnisse intraoperative DBT und Inter-Reader-Vergleich

In diesem Abschnitt werden die relevanten Ergebnisse der Auswertung der Bewertungsbögen zur DBT analog zum Vorgehen bei der intraoperativen PR (vgl. Abschnitt 3.3) dargestellt.

3.4.1 Präparatgröße

Für die Messung der Präparatgröße in der DBT wird aus denselben Gründen wie bei der PR die geringste Streuung zwischen den Readern erwartet (vgl. Abschnitt 3.3.1). Abbildung 3.11 zeigt den Vergleich der Messwertpaare beider Reader für die Präparatgröße in der DBT am Beispiel APx. Auf Basis des Korrelationsdiagramms (links) ist eine grundsätzlich gute Korrelation zwischen den Messungen der Reader zu erkennen. Ausreißer liegen zwar zu beiden Seiten der Winkelhalbierenden vor, wobei tendenziell mehr Werte oberhalb dieser liegen und somit durch tendenziell größere Messwerte von Reader 2 zu erklären sind. Die Steigung der Regressionsgerade hat den idealen Wert 1, das 95%-Konfidenzintervall umschließt den Wert 1 in engen Grenzen. Trotzdem kann aufgrund des positiven y-Achsenabschnitts nach Passing Bablok keine vollständige Übereinstimmung bzw. Austauschbarkeit festgestellt werden.

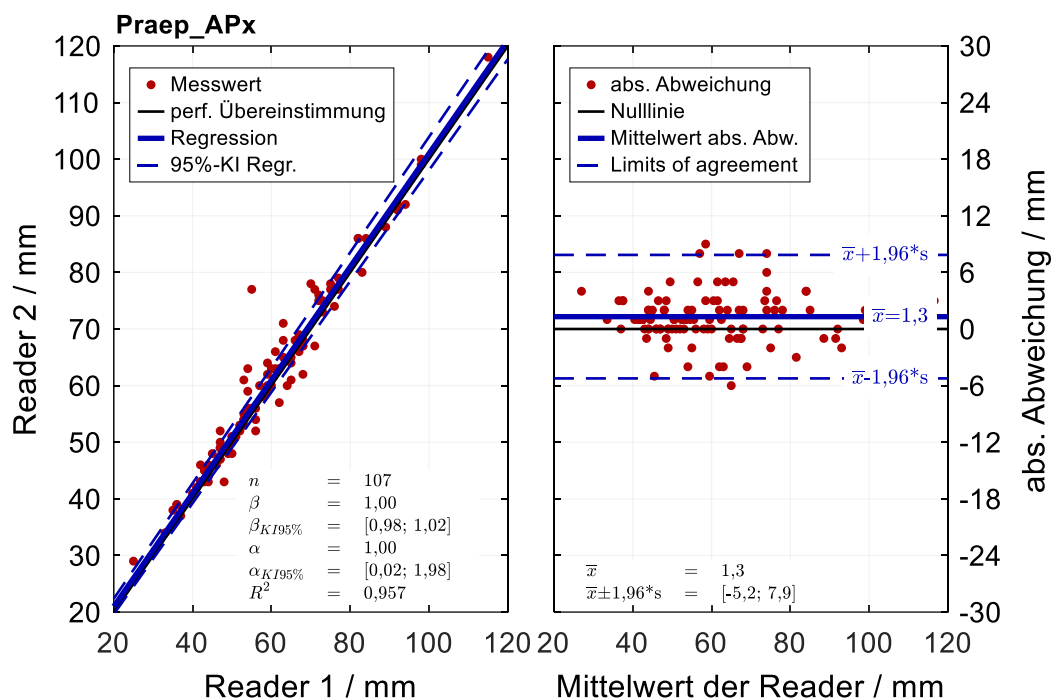


Abbildung 3.11: Inter-Reader-Vergleich der Größe des Präparats am Beispiel APx (DBT)

Die mittlere Abweichung zwischen den Messungen beider Reader liegt bei 1,3 mm. Die LoA, also die Intervallgrenzen innerhalb derer bei Annahme standardnormalverteilter Messdifferenzen 95 % der Messstreuung der Grundgesamtheit zu erwarten sind, liegen bei -5,2...7,9 mm. Im Vergleich zur PR und zum Messbereich von ca. 0-120 mm (PR: 0-140 mm) erscheint die 95%-Streubreite mit 13,1 mm als eher klein. Trotzdem liegt diese immer noch oberhalb des im Rahmen dieser Arbeit festgelegten Gütekriteriums von ± 5 mm.

Tabelle 3.10 zeigt die Vergleichsgrößen für die weiteren erhobenen Messgrößen.

Die jeweiligen Diagramme können im Anhang 8.6 eingesehen werden.

Die fast gleichbleibende mittlere Abweichung deutet auf einen systematischen Versatz zwischen den Readern hin. Die Messgrößen in den Ebenen APz und LLz verfehlen das Gütekriterium dieser Arbeit nur knapp. Die Analyse der Übereinstimmung auf Basis der weiteren Messgrößen des Herdbefundes bestätigt die oben beschriebenen Erkenntnisse des Inter-Reader-Vergleichs.

Tabelle 3.10: Inter-Reader-Vergleichswerte Größe des Präparats (DBT)

	Mittlere Abweichung	Standardabw. Differenzen	LoA	
	mm	mm	mm	
Praep_APx	1,3084	3,3376	-5,2334	7,8502
Praep_APz	1,2710	2,5125	-3,6535	6,1955
Praep_LLy	1,2523	5,0918	-8,7276	11,2322
Praep_LLz	1,3551	2,7239	-3,9838	6,6941

3.4.2 Größe Herdbefund

In Abbildung 3.12 ist der Inter-Reader-Vergleich der Größe des Herdbefundes am Beispiel der Messgröße APx dargestellt.

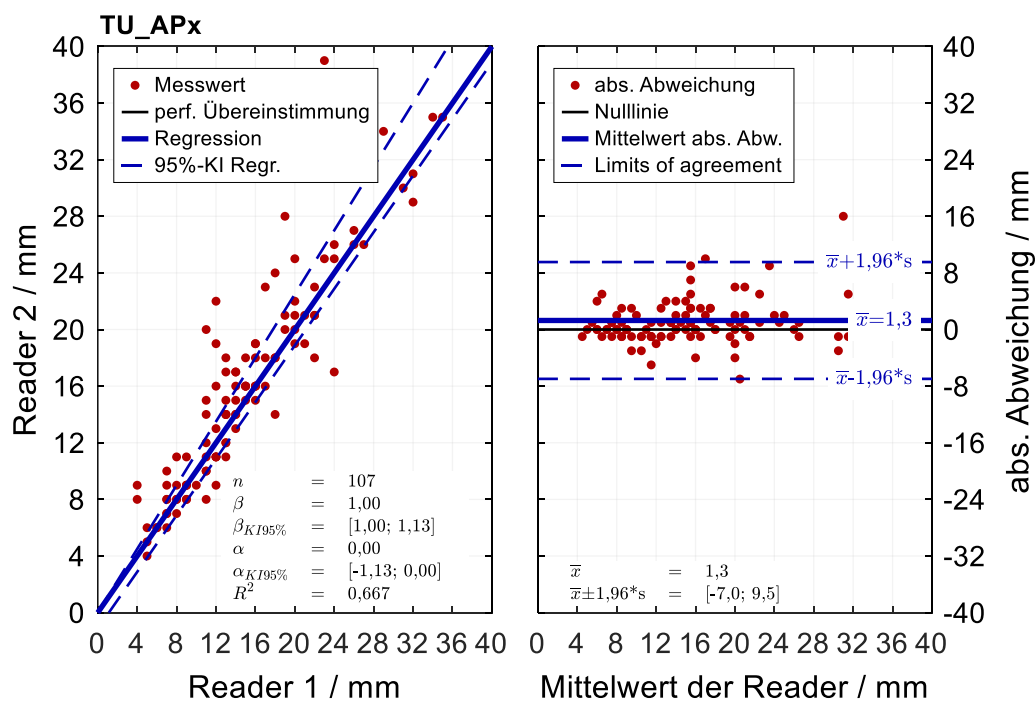


Abbildung 3.12: Inter-Reader-Vergleich der Größe des Herdbefundes am Beispiel APx (DBT)

Es ist eine grundsätzlich gute Korrelation zwischen den Messungen der Reader zu erkennen, wobei tendenziell mehr Ausreißer oberhalb der Winkelhalbierenden und somit bei größeren Messwerten von Reader 2 liegen. Gemäß der Methode nach Passing Bablok liegt formal eine vollständige Übereinstimmung vor. Die mittlere Abweichung zwischen den Messungen beider Reader liegt bei 1,3 mm.

Die LoA liegen bei -7,0...9,5 mm. Im Vergleich zum Messbereich von ca. 0-40 mm erscheint diese 95%-Streubreite mit 16,5 mm hoch und überschreitet das in dieser Arbeit festgelegte Gütekriterium. Es kann daher eine gute Korrelation mit formal vollständiger Übereinstimmung bei jedoch hoher Streuung festgehalten werden. Tabelle 3.11 zeigt die Vergleichswerte auch für die weiteren erhobenen Messgrößen. Die jeweiligen Diagramme können im Anhang 8.6 eingesehen werden.

Tabelle 3.11: Inter-Reader-Vergleichswerte Größe des Herdbefundes (DBT)

	Mittlere Abweichung	Standardabw. Differenzen	LoA	
	mm	mm	mm	
TU_APx	1,2804	4,2109	-6,9731	9,5338
TU_APz	1,6636	4,0304	-6,2361	9,5632
TU_LLy	1,4245	4,2245	-6,8555	9,7046
TU_LLz	1,7736	4,6132	-7,2683	10,8154

Die Analyse der weiteren Messgrößen bestätigt die oben beschriebenen Erkenntnisse des Inter-Reader-Vergleichs, einer beinahe konstant mittleren Abweichung hin zu größeren Messwerten von Reader 2, sowie tendenziell hoher Streuung.

3.4.3 Randabstand

Im Vergleich zu den anderen DBT-Messgrößen (Präparatgröße und Herdbefund) wird für die Messung des Randabstandes tendenziell eine größere Streuung aufgrund der kleinen Messwerte sowie der nachteiligen optischen Trennschärfe erwartet. In Abbildung 3.13 ist der Inter-Reader-Vergleich der Messungen des Randabstandes am Beispiel APkaudal zu sehen. Genau wie bei der PR kann auch hier ein geringerer Stichprobenumfang beobachtet werden. Auf Basis des Korrelationsdiagramms (links) ist eine grundsätzliche Korrelation zwischen den Messungen der Reader zu erkennen. Ausreißer liegen tendenziell bei größeren Werten von Reader 1 vor.

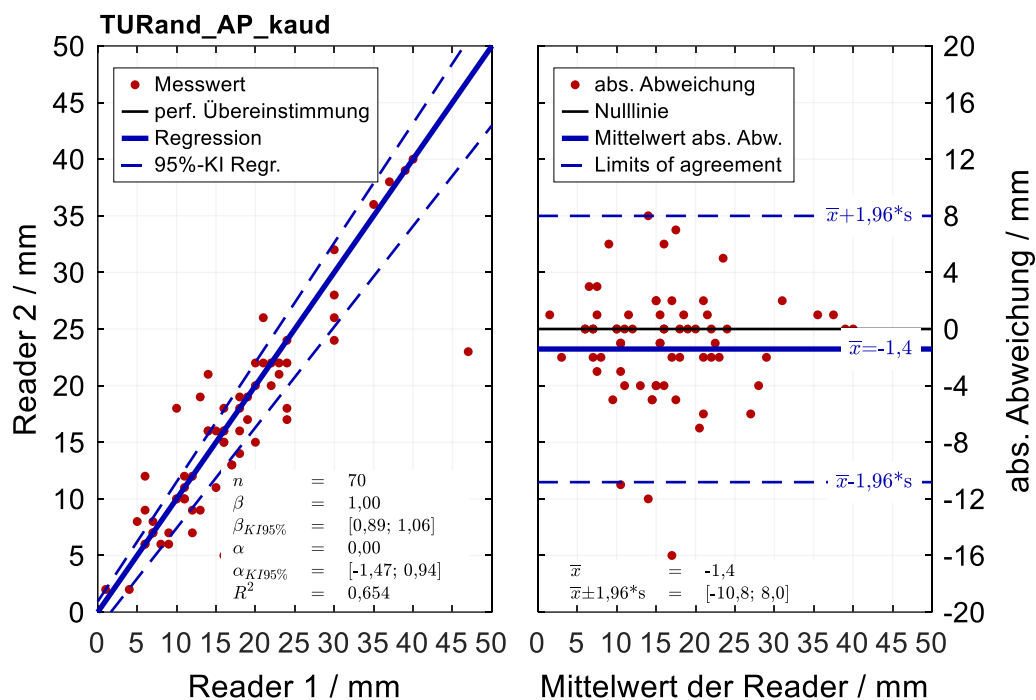


Abbildung 3.13: Inter-Reader-Vergleich des Randabstandes am Beispiel AP-kaudal (DBT)

Trotzdem kann gemäß der Methode nach Passing Bablok formal eine vollständige Übereinstimmung der Reader festgestellt werden. Die mittlere Abweichung zwischen den Messungen beider Reader liegt bei -1,4 mm. Die LoA liegen bei -10,8...8,0 mm. Im Vergleich zum Messbereich von ca. 0-50 mm erscheint diese 95%-Streuung mit 18,8 mm hoch und verletzt zudem das mit Blick auf die klinische Relevanz festgelegte Gütekriterium dieser Arbeit in Höhe von ± 5 mm. Es kann daher eine gute Korrelation mit formal vollständiger Übereinstimmung bei jedoch hoher Streuung festgehalten werden.

Tabelle 3.12 stellt die Vergleichswerte für die weiteren erhobenen Messgrößen gegenüber. Die jeweils zugehörigen Diagramme können im Anhang 8.6 eingesehen werden. Die Analyse der Übereinstimmung auf Basis der weiteren Messgrößen des Randabstandes bestätigt die oben beschriebenen Erkenntnisse des Inter-Reader-Vergleichs. Es besteht eine geringe mittlere Abweichung zwischen den Readern bei gleichzeitig hoher Streuung.

Tabelle 3.12: Inter-Reader-Vergleichswerte Größe des Randabstandes (DBT)

	n	Mittlere Abweichung	Standardabw. Differenzen	LoA	
	-	mm	mm	mm	
AP_mam	96	-0,3333	9,7091	-19,3632	18,6965
AP_per	100	0,1300	8,5619	-16,6513	16,9113
AP_kran	71	-0,0282	5,7246	-11,2483	11,1920
AP_kaud	70	-1,4143	4,7988	-10,8200	7,9914
AP_med	20	0	4,0131	-7,8657	7,8657
AP_lat	20	-0,2500	4,0246	-8,1382	7,6382
LL_dors	100	-0,7400	4,2607	-9,0911	7,6111
LL_kut	101	-0,4851	4,4127	-9,1341	8,1638
LL_mam	96	-1,5729	8,9722	-19,1584	16,0125
LL_per	101	-0,3366	7,8272	-15,6780	15,0047

3.4.4 Inter-Reader-Vergleich der Diagnose R0/R1

Der Vergleich der aus den Randabstandsmessungen abgeleiteten Diagnosen der beiden Reader ist in Tabelle 3.13 dargestellt.

Tabelle 3.13: Kontingenztabelle R0/R1-Vergleich intraoperative DBT

Kontingenztabelle intraoperative DBT R0/R1 (abs.)

		Reader 1		Randhäufigkeit
		R0	R1	
Reader 2	R0	44	5	49
	R1	10	48	58
Randhäufigkeit		54	53	107

Kontingenztabelle intraoperative DBT R0/R1 (rel.)

		Reader 1		Randhäufigkeit
		R0	R1	
Reader 2	R0	0,41121	0,046729	0,45794
	R1	0,093458	0,4486	0,54206
Randhäufigkeit		0,50467	0,49533	

p0 0,85981

pe 0,49961

K2 0,71985

unterer Wert 0,58838

oberer Wert 0,85131

Es besteht eine übereinstimmende Beurteilung beider Reader bei R0/R1 in 92 von 107 gemeinsam bearbeiteten Fällen aus Datensatz 1 (86 %). Die Inter-Reader-Reliabilität ($K2 = 0,72$) kann gemäß Altmann [86] als gut eingestuft werden. Das 95%-Konfidenzintervall liegt im Bereich gut bis sehr gut. Es liegt somit eine gute Übereinstimmung der Reader in der Urteilsausprägung trotz hoher Streuungen in der Messung der Randabstände vor.

3.5 Vergleich der PR und DBT

In diesem Abschnitt werden die Ergebnisse des Inter-Methoden Vergleichs zwischen PR und DBT vorgestellt. Hierzu werden zum einen über die Reader aggregierte Ergebnisgrößen innerhalb einer Methode gegenübergestellt sowie zum anderen die R0/R1-Diagnose von dem jeweils selben Reader zwischen den Methoden verglichen.

3.5.1 Vergleich von PR und DBT anhand aggregierter Ergebnisgrößen des Inter-Reader-Vergleichs

Um einen übersichtlichen und aussagekräftigen Vergleich zwischen den Methoden zu ermöglichen, werden im Rahmen dieser Arbeit über alle Messgrößen arithmetisch gemittelte Vergleichswerte aus dem Inter-Reader-Vergleich verwendet. Hierbei wird zum einen die mittlere Abweichung zwischen den Readern sowie als Maß für die Streuung der Messwerte zwischen den Readern die Standardabweichung verwendet. Da nur eine qualitative, vergleichende Aussage zwischen den Methoden abgeleitet werden soll, ist es ausreichend die Standardabweichung selbst und nicht die LoA heranzuziehen. Zum Vergleich der Inter-Reader Reliabilität in der R0/R1-Diagnose wird als Vergleichswert Cohens Kappa verwendet.

Tabelle 3.14 zeigt die aggregierten Vergleichswerte für die Präparatgröße, für die Herdbefundgröße und für die Randabstandgröße der DBT und PR sowie für Cohens Kappa. Im Vergleich der mittleren Inter-Reader-Abweichung können nur kleine Unterschiede zwischen den beiden Verfahren PR und DBT identifiziert werden. Es treten im Einklang der Analyse der Ergebnisse auf Basis einzelner

Tabelle 3.14: Gemittelte Vergleichswerte zur Beurteilung der Inter-Reader Reliabilität der intraoperativen DBT und PR

			DBT	PR
über alle Präparat- Größen	Mittlere Abweichung	mm	1,2967	0,9773
	Standard- abweichung	mm	3,4165	7,0374
über alle Herdbefund- Größen	Mittlere Abweichung	mm	1,5355	1,1745
	Standard- abweichung	mm	4,2698	5,8773
über alle Randabstand- Größen	Mittlere Abweichung	mm	-0,5030	-0,3431
	Standard- abweichung	mm	6,2305	6,6570
	Cohens Kappa	-	0,71985	0,56897

Messgrößen in Abschnitt 3.3 und 3.4 auch im Mittel über alle Fälle und Messgrößen in der PR und der DBT nur kleine Abweichungen zwischen den Readern auf. Hierbei ist die PR minimal im Vorteil, wenngleich der Unterschied so klein ausfällt, dass er aus klinischer Sicht nicht relevant erscheint.

Die Streuung der Messwerte zwischen den Readern ist in der DBT bei der Größe des Präparats und des Herdbefundes deutlich kleiner als in der PR. Bei der Messung des Randabstandes liegen beide Verfahren auf einem ähnlichen Niveau. Trotzdem zeigt die DBT eine deutlich höhere Inter-Reader Reliabilität mit Blick auf die Diagnose von R0/R1 als die PR. Aus dieser Beobachtung kann die Hypothese abgeleitet werden, dass die DBT durch die größere Anzahl an Bildern/Ebenen eine verbesserte Trennschärfe und damit eine bessere Interpretierbarkeit ermöglicht. Diese Hypothese wird von der subjektiven Einschätzung der Reader untermauert, die im Rahmen des Beurteilungsbogens für jeden Fall erhoben wurde. In Abbildung 3.14 ist die Verteilung der Antworten beider Reader auf die Frage wie gut der Herdbefund messbar war für Datensatz 1 dargestellt. Es ist ersichtlich, dass die DBT eine subjektiv bessere Messbarkeit des Herdbefundes ermöglicht. Dies lässt sich auch am Unterschied der gewichteten Mittelwerte der Antworten ableiten, der unter Zuordnung von Schulnoten 1 bis 5,

zu den Antwortmöglichkeiten gebildet werden kann. Hier weist die DBT im Schnitt mit 2,36 ein positiveres Bild auf als die PR mit 2,6.

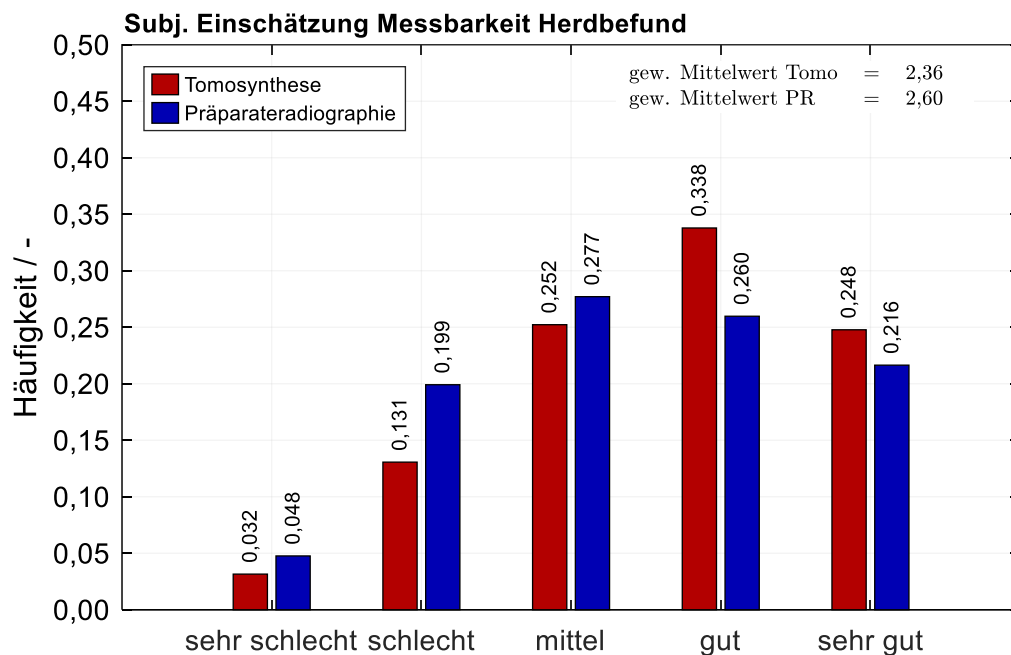


Abbildung 3.14: Subjektive Einschätzung der Reader auf die Frage wie gut der Herdbefund in DBT und PR messbar ist

3.5.2 Inter-Methoden Vergleich der Diagnose R0/R1

Um auf Basis der R0/R1-Diagnose einen Inter-Methoden Vergleich zwischen DBT und PR zu ermöglichen, wurden die Urteilsausprägungen in den unterschiedlichen Methoden für beide Reader separat in je einer Kontingenztafel dargestellt. Tabelle 3.15 stellt das Ergebnis des Vergleichs für Reader 1 dar. Das 95%-Konfidenzintervall liegt im hinreichenden bis guten Bereich. Es kann daher festgehalten werden, dass derselbe Reader in den beiden Methoden nicht immer zum gleichen Ergebnis kommt und daher faktisch Unterschiede in der diagnostischen Güte der Verfahren bestehen.

In Tabelle 3.16 wird in gleicher Weise das Ergebnis des Vergleichs zwischen den Methoden für Reader 2 dargestellt.

Tabelle 3.15: Kontingenztabelle zur Inter-Methoden-Reliabilität R0/R1 zwischen intraoperativer PR und DBT (Reader 1)

Kontingenztabelle intraoperative PR vs DBT Reader 1 R0/R1 (abs.)

		DBT		Randhäufigkeit
		R0	R1	
PR	R0	40	13	53
	R1	14	40	54
Randhäufigkeit		54	53	107

Kontingenztabelle intraoperative PR vs DBT Reader 1 R0/R1 (rel.)

		DBT		Randhäufigkeit
		R0	R1	
PR	R0	0,37383	0,1215	0,49533
	R1	0,13084	0,37383	0,50467
Randhäufigkeit		0,50467	0,49533	

p0 0,74766

pe 0,49996

K2 0,49537

unterer Wert 0,33079

oberer Wert 0,65996

Tabelle 3.16: Kontingenztabelle zur Inter-Methoden-Reliabilität R0/R1 zwischen intraoperativer PR und DBT (Reader 2)

Kontingenztabelle intraoperative PR vs DBT Reader 2 R0/R1 (abs.)

		DBT		Randhäufigkeit
		R0	R1	
PR	R0	36	20	56
	R1	13	38	51
Randhäufigkeit		49	58	107

Kontingenztabelle intraoperative PR vs DBT Reader 2 R0/R1 (rel.)

		DBT		Randhäufigkeit
		R0	R1	
PR	R0	0,33645	0,18692	0,52336
	R1	0,1215	0,35514	0,47664
Randhäufigkeit		0,45794	0,54206	

p0 0,69159

pe 0,49803

K2 0,38559

unterer Wert 0,21126

oberer Wert 0,55992

In 74 von 107 bearbeiteten Fällen (69,2 %) kommt Reader 2 in PR und DBT zum gleichen Ergebnis. Die Inter-Methoden Reliabilität auf Basis von Cohens Kappa ($K_2 = 0,39$) muss gemäß Altman [86] als hinreichend eingestuft werden. Das 95%-Konfidenzintervall liegt im hinreichenden bis moderaten Bereich.

Folglich zeigt sich für Reader 2 eine etwas schwächere Inter-Methoden Reliabilität als bei Reader 1. Neben den bereits für Reader 1 beschriebenen faktischen Unterschieden zwischen der diagnostischen Güte von PR und DBT kann auch festgehalten werden, dass der Reader einen Einfluss auf die Inter-Methoden Reliabilität hat.

3.6 Vergleich des Goldstandards (Histologie) mit PR und DBT

Um den in Abschnitt 3.5 festgestellten Unterschied in der diagnostischen Güte von PR und DBT einordnen zu können und mögliche Vorteile für ein Verfahren zu bewerten, werden in diesem Abschnitt die Ergebnisse der radiologischen Verfahren mit den Ergebnissen des Goldstandards, der Histologie, verglichen.

Bei dem Vergleich der Ergebnisse aus diesen unterschiedlichen Methoden gilt es jedoch, die in der nachfolgenden Abbildung 3.15 festgehaltenen Herausforderungen zu lösen:

Herausforderungen 	Vorgehensweise 
1 Pathologin misst Präparat im Gegensatz zu radiologischen Verfahren nicht im eingespannten Zustand	Vergleich der Präparatgröße nicht sinnvoll , da durch Kompression des gesunden Gewebes Verzerrungen auftreten können
2 In Histologie wird lediglich die längste, kürzeste und mittlere Abmessung des Herdbefundes angegeben	Vergleich des jeweils maximalen Messwerts pro Fall und Methode/Reader mit dem Messwert der Histologie
3 In Histologie werden die Randabstände größer 10 mm nicht mehr mit expliziten Zahlenwerten , sondern mit >10 mm dokumentiert	Verminderter Stichprobenumfang schränkt Vergleich von Messwerten ein . Der Vergleich muss auf die Befundung R0/R1 gestützt werden
4 Histologische Kriterien für R0/R1 sind abweichend zu den in dieser Arbeit angelegten Kriterien (10 mm / 2 mm) für die Radiologie	Neuauswertung der R0/R1 Befunde entsprechend der in Histologie angegebenen, kleinsten Randabstände mit den in dieser Arbeit angelegten Kriterien

Abbildung 3.15: Herausforderungen beim Vergleich der Ergebnisse aus der Histologie und den radiologischen Verfahren

Aufgrund des Umstandes, dass die Pathologin das Präparat im Vergleich zum radiologischen Verfahren nicht im eingespannten Zustand misst, kann kein sinnvoller Vergleich der Präparatgröße erfolgen, da die Kompression des gesunden

Gewebes zu Verzerrungen führen könnte. Aus diesem Grund wird im Rahmen der vorliegenden Arbeit – unter der Annahme einer geringen Kompressibilität des Tumorgewebes – die Herdbefundgröße als Vergleichswert herangezogen. Eine zusätzliche Herausforderung besteht darin, dass in den vorliegenden Histologien jeweils nur die längste, kürzeste und mittlere Abmessung des Herdbefundes bestimmt worden ist, diese jedoch nicht im Koordinatensystem zugeordnet wurde. Um dennoch eine Vergleichbarkeit zu erreichen wird daher im Rahmen dieser Arbeit, die jeweils längste gemessene Herdbefundgröße aus den radiologischen Verfahren mit der Histologie verglichen. Darüber hinaus besteht die Herausforderung, dass die Histologie Randabstände zwar im Koordinatensystem orientiert angibt, diese aber zur Vereinfachung nur bei Randabständen kleiner 10 mm bestimmt. Dies schränkt die Anzahl der vergleichbaren Messwertpaare ein. Daher wird im Rahmen der vorliegenden Arbeit die Aussage zum Vergleich ausschließlich auf die R0/R1-Diagnose gestützt. Eine beispielhafte Darstellung des Vergleichs der Messwertpaare für die Randabstände ist der Vollständigkeit halber in Abschnitt 3.6.2 dargestellt. Aus Gründen der Lesbarkeit, sind in den nachfolgenden Kapiteln nur die Diagramme und Kontingenztafeln von Reader 1 dargestellt. Die zugehörigen Grafiken und Tabellen für Reader 2 sind im Anhang 8.7 zu finden. Um die R0/R1-Diagnose aus der Histologie und der Radiologie vergleichbar zu machen, wurde die R0/R1-Diagnose der Histologie wie in Abschnitt 2.3 beschrieben an die im Rahmen dieser Arbeit definierten Randabstandskriterien angepasst. Abweichend zu diesem Vorgehen ist in der klinischen Praxis die Vorhersage der finalen, histologischen Diagnose von Interesse. Hierzu ist ein geeignetes Randabstandskriterium zu definieren. Auf Basis der Ergebnisse nach o.g. Vorgehen, wird in Abschnitt 3.6.4 mittels Sensitivitätsanalyse untersucht, bei welchem Randabstandskriterium die beste Übereinstimmung in Bezug auf Sensitivität und Spezifität vorliegt.

3.6.1 *Vergleich der maximalen Herdbefundgröße*

Zum Inter-Methoden Vergleich zwischen dem Goldstandard Histologie und den radiologischen Verfahren auf Basis der maximalen Herdbefundgröße wird ebenfalls die Methodik nach Passing Bablok und Bland-Altman genutzt. In Abbildung

3.16 sind die entsprechenden Messwertpaare für den größten Messwert der Herdbefundgröße aus Histologie und PR für Reader 1 dargestellt.

Auf Basis des Korrelationsdiagramms ist eine grundsätzliche Korrelation zwischen den Messwerten aus Histologie und PR zu erkennen. Ausreißer liegen tendenziell bei größeren Werten von Reader 1, wobei ein Trend hin zu vermehrt größeren Abweichungen bei größeren Messwerten sichtbar ist. Dies spiegelt sich auch in der Steigung der Regressionsgerade wider, die größer eins ist. Eine vollständige Übereinstimmung liegt damit formal nicht vor. Der im Bland-Altman-Diagramm ersichtliche mittlere Fehler von 1,6 mm bestätigt die gegenüber der Histologie tendenziell zu großen Messwerte von Reader 1. Die LoA liegen bei -12,0...15,2 mm und die resultierende 95%-Streubreite erscheint im Vergleich zum Messbereich von ca. 0-60 mm groß. Die LoA überschreiten das im Rahmen der vorliegenden Arbeit festgelegte klinisch relevante Gütekriterium von ± 5 mm. Genau wie bei Reader 1 lassen sich auch bei Reader 2 dieselben Tendenzen feststellen (vgl. Abbildung 8.40 im Anhang). So ergeben sich im Vergleich zur Histologie ebenfalls tendenziell Abweichungen hin zu größeren Messwerten bei

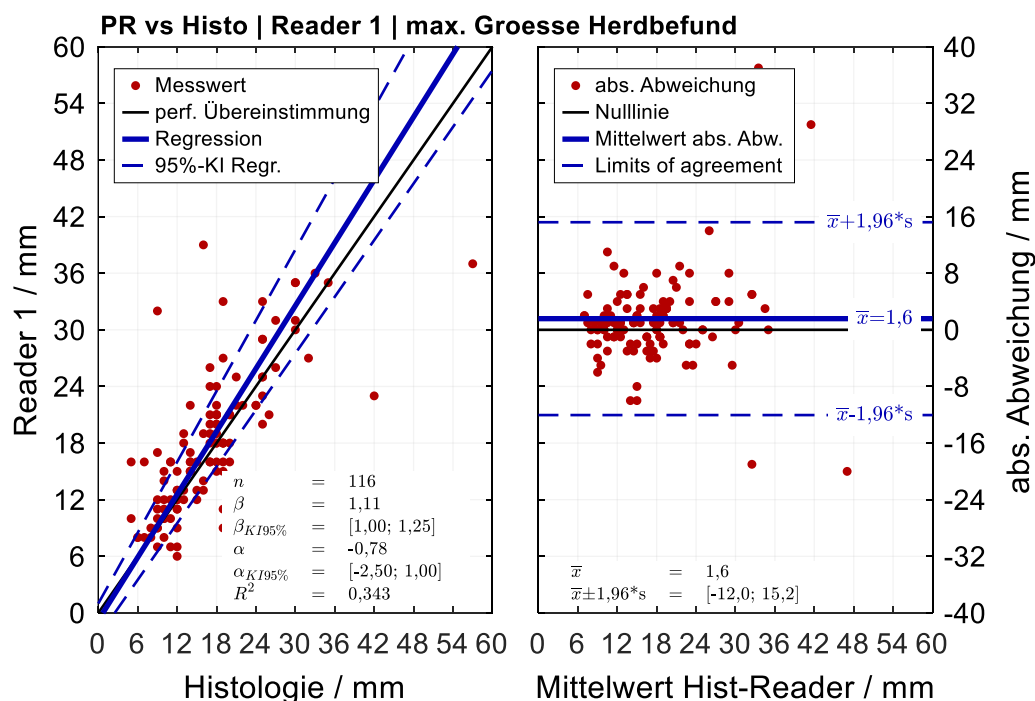


Abbildung 3.16: Vergleich der maximalen Größe des Herdbefundes zwischen PR und Goldstandard (Reader 1)

Reader 2 sowie eine positive Steigung der Regressionsgeraden. Auch hier kann keine vollständige Übereinstimmung festgestellt werden. Die LoA überschreiten ebenfalls das Gütekriterium.

Der Vergleich zwischen Goldstandard und DBT für Reader 1 ist in Abbildung 3.17 dargestellt.

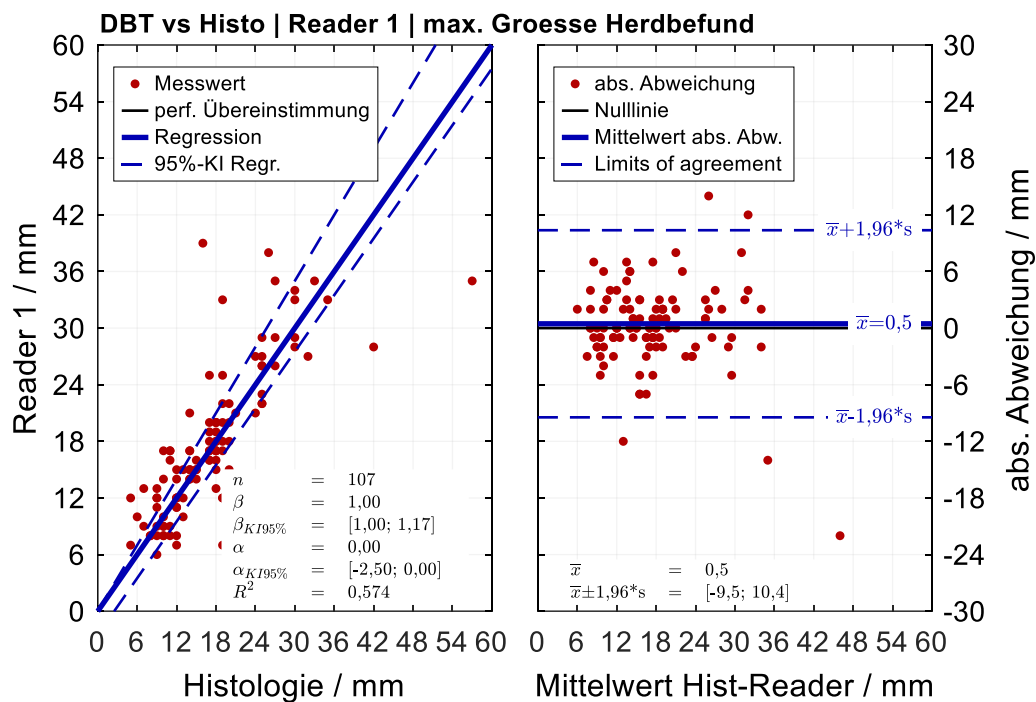


Abbildung 3.17: Vergleich der maximalen Größe des Herdbefundes zwischen DBT und Goldstandard (Reader 1)

Auf Basis des Korrelationsdiagramms ist eine grundsätzliche Korrelation zwischen den Messwerten aus Histologie und DBT zu erkennen. Ausreißer liegen tendenziell bei größeren Werten von Reader 1 im Vergleich zur Histologie. Dies ist an dem einseitig aufgeweiteten Konfidenzintervallen der Regressionsgerade zu erkennen. Die Steigung der Regressionsgerade liegt jedoch bei 1 und der y-Achsenabschnitt bei 0, so dass nach Passing Bablok formal eine vollständige Übereinstimmung besteht. Die mittlere Abweichung über alle Messwertpaare hinweg liegt bei 0,5 mm, die LoA bei -9,5...10,4 mm.

Die im Anhang dargestellte Abbildung 8.41 zeigt den Vergleich der Histologie mit der DBT für Reader 2. Im Korrelationsdiagramm ist ersichtlich, dass trotz grundsätzlicher Korrelation im Vergleich zur Histologie tendenziell größere Messwerte in der DBT von Reader 2 ermittelt wurden. Dieser Trend verstärkt sich hin zu

größeren Messwerten weiter, sodass eine Steigung der Regressionsgerade resultiert, die größer eins ist, wohingegen der y-Achsenabschnitt negativ ist. Eine formale Übereinstimmung nach Passing Bablok kann hier nicht abgeleitet werden. Im Bland-Altman-Diagramm ist die mittlere Abweichung von 2,4 mm sowie LoA von -8,6...13,4 mm ersichtlich.

Tabelle 3.17 stellt die Vergleichswerte der mittleren Abweichung, der Standardabweichung der Differenzen als Maß für die Streuung sowie die LoA zur Einordnung der klinischen Güte für beide Verfahren und beide Reader gegenüber.

Tabelle 3.17: Vergleichswerte für die max. Größe des Herdbefundes für PR und DBT im Vergleich zur Histologie

	Mittlere Abweichung		Standardabw. Differenzen		LoA	
	Reader 1	Reader 2	Reader 1	Reader 2	Reader 1	Reader 2
	mm	mm	mm	mm	mm	mm
PR vs. Histo	1,5776	2,6552	6,9509	5,8072	-12,0463 15,2014	-8,7269 14,0373
DBT vs. Histo	0,4579	2,3645	5,0587	5,6140	-9,4571 10,3729	-8,6389 13,3679

Bei der Vergleichsgröße der mittleren Abweichungen zeigen sich gegenüber dem Goldstandard Histologie bei beiden Readern moderate Abweichungen. Tendenziell wird in beiden Verfahren zu groß gemessen, wobei die mittlere Abweichung in der DBT kleiner ausfällt als in der PR. Die Streuung der Messwerte aus der DBT ist im Vergleich zur Streuung der Messwerte aus der PR für beide Reader kleiner. Beides deutet im Einklang mit den Ergebnissen aus dem Inter-Reader-Vergleich auf ein tendenziell robusteres radiologisches Verfahren der DBT hin. Die LoA sind entsprechend der mittleren Abweichung und der Standardabweichung bei der DBT ebenfalls kleiner, liegen jedoch bei beiden radiologischen Verfahren außerhalb des im Rahmen dieser Arbeit aus Sicht der klinischen Relevanz definierten Gütekriteriums von ± 5 mm. Insgesamt lässt sich jedoch

festhalten, dass die DBT zur Messung des Herdbefundes relativ zur PR besser geeignet ist.

3.6.2 Vergleich der Randabstandsmessung

In Abbildung 3.18 wird exemplarisch der Goldstandard Histologie mit dem Verfahren PR für den Reader 1 in Bezug auf die Messung des Randabstandes verglichen.

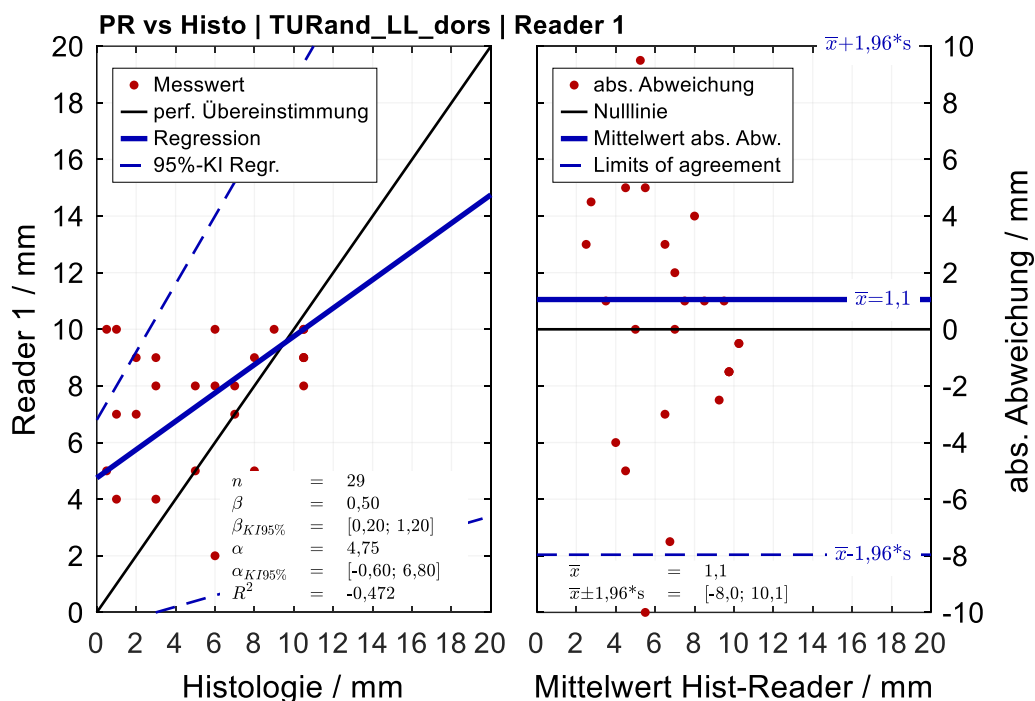


Abbildung 3.18: Vergleich des Randabstandes LLdorsal zwischen PR und Goldstandard (Reader 1)

Wie einführend beschrieben, schränkt die Vorgehensweise in der Histologie (nur explizite numerische Dokumentation von Randabständen <10 mm) den Stichprobenumfang ein. Für den Vergleich wurde der Randabstand LLdorsal beispielhaft für einen Randabstand mit vergleichsweise hohem Stichprobenumfang ($n = 29$) gewählt. Im Korrelationsdiagramm (links) ist eine hohe Streuung der Messwerte zu erkennen. Die Regressionsgerade hat eine Steigung von 0,5 und es liegen sehr große Konfidenzintervalle vor. Das negative R^2 deutet darauf hin, dass die Regression die Streuung schlechter erklärt als eine horizontale Linie und muss daher als nicht aussagekräftig eingestuft werden. Die Messwerte von Reader 1

liegen tendenziell oberhalb der Randabstandswerte aus der Histologie. Die mittlere Abweichung liegt bei 1,1 mm. Die LoA liegen bei -8,0...10,1 mm. Im Vergleich zum Messbereich von ca. 0-20 mm erscheint diese 95%-Streubreite mit 18,1 mm sehr hoch. Die Streubreite ist für eine robuste Beurteilung von R0/R1 auf Basis des festgelegten Gütekriterium von ± 5 mm zu hoch.

Die im Anhang befindliche Abbildung 8.42 stellt den Vergleich für Reader 2 dar. Es ergibt sich ein Stichprobenumfang von $n = 24$. Genau wie bei Reader 1 ist im Korrelationsdiagramm eine hohe Streuung erkennbar, die nicht durch die Regressionsgerade erklärt werden kann (negatives R^2). Die mittlere Abweichung zwischen den Messungen von Reader 2 mit dem Verfahren PR und der Histologie als Goldstandard liegt bei 1,6 mm. Die LoA liegen bei -6,3...9,6 mm. Im Vergleich zum Messbereich von ca. 0-20 mm erscheint diese 95%-Streubreite mit 15,9 mm sehr hoch. Genau wie bei Reader 1 ist die Streubreite für eine robuste Beurteilung von R0/R1 auf Basis des festgelegten Gütekriterium von ± 5 mm zu hoch.

Ein ähnliches Bild ergibt sich beim Vergleich der Messwerte für Reader 1 und 2 für die DBT. Die Vergleiche für die Messgröße Randabstand LLdorsal ist in Abbildung 3.19 für Reader 1 und in Abbildung 8.43 (im Anhang) für Reader 2 dargestellt. Auch hier ergibt sich das gleiche Bild einer hohen Streuung, die nicht durch die Regression erklärt werden kann. Wie zuvor bei der PR sind die Streubreite bzw. die LoA für eine robuste Beurteilung von R0/R1 auf Basis des festgelegten Gütekriterium von ± 5 mm zu hoch.

Die mittleren Abweichungen sowie die zugehörigen Standardabweichungen und die Stichprobenumfänge für die einzelnen Randabstandsmesswerte sind in Tabelle 3.18 für die PR und in Tabelle 3.19 für die DBT zusammengefasst.

Es ist ersichtlich, dass die für LL-dorsal dargestellten Ergebnisse auf den höchsten Stichprobenumfängen basieren. Bei den anderen Randabständen ergeben sich zum Teil deutlich geringere Stichprobenumfänge. Es fällt auf, dass die Stichprobenumfänge der Reader sich für einzelne Randabstände (insb. In der AP-Ebene) unterscheiden. Neben Einflussfaktoren wie nicht eindeutig erkennbaren Randabständen, bei denen ein Reader eine Messung ableitet, der andere jedoch nicht, könnte dies auch darauf zurückzuführen sein, dass die Reader die

minimalen Randabstände in anderen Ebenen identifizieren. Aufgrund der kleinen Stichprobenumfänge sind statistisch relevante Aussagen jedoch nicht abzuleiten.

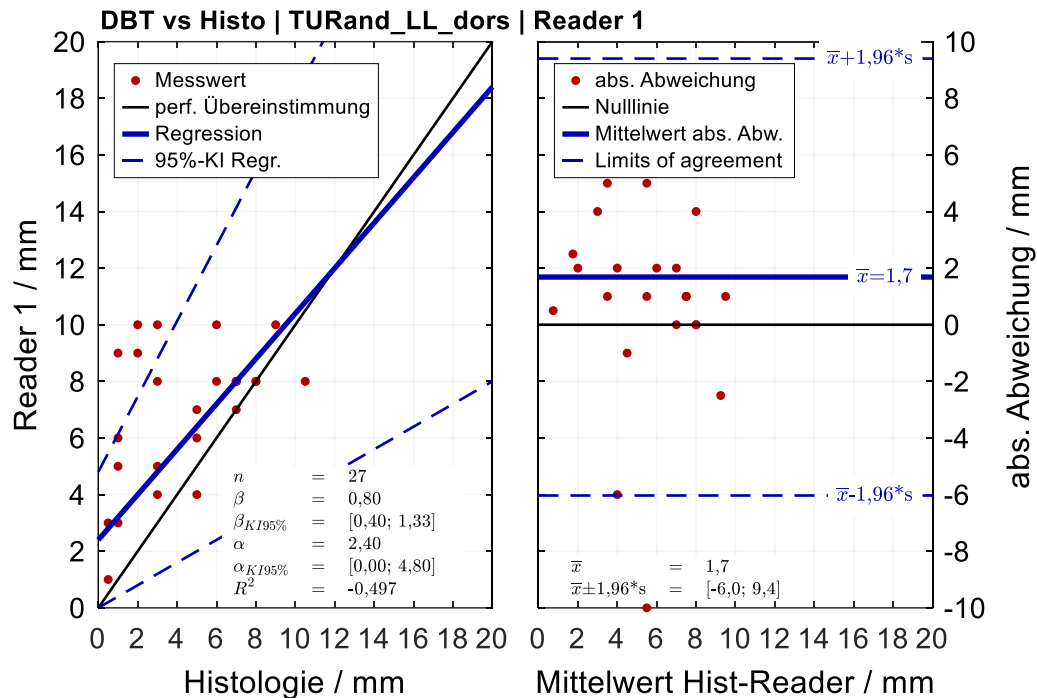


Abbildung 3.19: Vergleich des Randabstandes LLdorsal zwischen DBT und Goldstandard (Reader 1)

Tabelle 3.18: Vergleich der Randabstandsmessung von Reader 1 und Reader 2 mit der Histologie (PR)

	Reader 1			Reader 2		
	n	mit. Abweichung	Standard-abw.	n	mit. Abweichung	Standard-abw.
	-	mm	mm	-	mm	mm
AP_mam	9	-2,5000	3,7249	7	-1,8571	2,1931
AP_per	6	1,0833	3,4120	7	1,2857	3,6499
AP_kran	18	-0,0556	5,2493	17	1,3824	5,6583
AP_kaud	16	0,7500	3,8297	13	0,8846	4,7353
AP_med	6	1,5833	3,1212	8	1,0000	3,5757
AP_lat	4	5,5000	1,9149	8	1,8125	3,0111
LL_dors	29	1,0517	4,5988	24	1,6458	4,0553
LL_kut	24	1,5000	4,2440	21	1,1905	4,2440
LL_mam	6	-2,5833	2,6157	6	-1,5833	1,9600
LL_per	4	2,5000	3,4881	10	1,1500	3,0736

Tabelle 3.19: Vergleich der Randabstandsmessung von Reader 1 und Reader 2 mit der Histologie (DBT)

	Reader 1			Reader 2		
	n	mit. Abwei- chung	Standard- abw.	n	mit. Abwei- chung	Standard- abw.
	-	mm	mm	-	mm	mm
AP_mam	12	-0,5000	1,9540	8	-1,4375	2,9208
AP_per	5	2,7000	3,0537	7	0,6429	3,6596
AP_kran	23	1,4130	5,0083	21	0,6905	4,5647
AP_kaud	17	1,9118	3,5979	21	1,6667	3,5508
AP_med	5	3,7000	4,2071	8	2,6250	5,8355
AP_lat	5	2,7000	3,7683	6	2,5833	2,9055
LL_dors	27	1,6852	3,9374	28	0,3929	3,9846
LL_kut	16	0,9063	4,3636	18	1,1944	4,4296
LL_mam	7	-0,8571	2,7343	11	-3,1818	2,8484
LL_per	7	1,0714	2,0295	7	-0,4286	3,2071

3.6.3 Vergleich der R0/R1-Diagnose

Zum Inter-Methoden Vergleich in Bezug auf die R0/R1-Diagnose zwischen dem Goldstandard Histologie und den radiologischen Verfahren werden Kontingenztafeln verwendet. Zusätzlich zu den beim Inter-Reader-Vergleich verwendeten Größen wird die Sensitivität sowie die Spezifität als zentrale Vergleichsgröße für die radiologischen Ergebnisse gegenüber der Histologie genutzt. Darüber hinaus wird abschließend analysiert, ob eine Abhängigkeit der Übereinstimmung der Diagnose von der Herdbefundgröße vorliegt.

Vergleich PR und Histologie

Die Kontingenztabelle Tabelle 3.20 zeigt den Vergleich der R0/R1-Diagnose des Reader 1 mit dem Verfahren PR zur Histologie. Zwischen den Diagnosen von Reader 1 mit der PR und dem histologischen Befunden liegen in 69 von 116 Fällen (59,5 %) eine übereinstimmende Beurteilung vor. Die Sensitivität, also der Anteil der aus Sicht der Histologie durch das radiologische Verfahren richtig identifizierten R1-Fälle, liegt bei 55,4 %. Die Spezifität, die den Anteil der durch das radiologische Verfahren aus Sicht der Histologie richtig identifizierten R0 Befunde angibt, liegt bei 86,7 %.

Tabelle 3.20: Kontingenztabelle R0/R1-Vergleich PR und Histologie (Reader 1)*Kontingenztabelle intraoperative PR vs Histologie Reader 1 R0/R1 (abs.)*

		PR Reader 1			Rand- häufigkeit		
		R0	R1				
Histo	R0	13	2	15	Spezifität	0,867	
	R1	45	56	101	Sensitivität	0,554	
Rand- häufigkeit		58	58	116			

Kontingenztabelle intraoperative PR vs Histologie Reader 1 R0/R1 (rel.)

		PR Reader 1			Rand- häufigkeit
		R0	R1		
Histo	R0	0,11207	0,017241	0,12931	
	R1	0,38793	0,48276	0,87069	
Rand- häufigkeit		0,5	0,5		

p0	0,59483		
pe	0,5		
K2	0,18966		
unterer Wert	0,01098	oberer Wert	0,36833

Die Inter-Methoden-Reliabilität auf Basis von Cohens Kappa $K2 = 0,19$ muss gemäß Altman [86] als knapp nicht ausreichend eingestuft werden. Das 95%-Konfidenzintervall liegt im nicht ausreichend bis hinreichenden Bereich. Insgesamt muss eine nicht ausreichende Urteils-übereinstimmung festgestellt werden. Um mögliche Einflüsse aus subjektiver Interpretation der Reader weitestgehend zu eliminieren und einen möglichst sauberen Vergleich der Methoden zu ermöglichen, sollen die oben beschriebenen Ergebnisse aus Datensatz 1 mit Datensatz 2 (eindeutige Fälle, vgl. Kapitel 2.1) verifiziert werden. Die resultierende Kontingenztabelle ist in Tabelle 3.21 dargestellt. Unter Verwendung des zweiten Datensatzes zeigt sich ein gesteigerter Anteil der übereinstimmenden Beurteilung bezüglich R0/R1 in 24 von 32 Fällen (75 %).

Tabelle 3.21: Kontingenztafel R0/R1-Vergleich PR und Histologie (Reader 1, Datensatz 2)

Kontingenztafel intraoperative PR vs Histologie Reader 1 R0/R1 (abs.)

		PR Reader 1				
		R0	R1	Rand-häufigkeit		
Histo	R0	13	0	13	Spezifität	1,000
	R1	8	11	19	Sensitivität	0,579
Rand-häufigkeit		21	11	32		

Kontingenztafel intraoperative PR vs Histologie Reader 1 R0/R1 (rel.)

		PR Reader 1				
		R0	R1	Rand-häufigkeit		
Histo	R0	0,40625	0	0,40625		
	R1	0,25	0,34375	0,59375		
Rand-häufigkeit		0,65625	0,34375			

p0 0,75

pe 0,4707

K2 0,52768

unterer Wert 0,24423 **oberer Wert** 0,81112

Die entsprechend gestiegene Intermethoden-Reliabilität auf Basis von Cohens Kappa ($K2 = 0,528$) kann gemäß Altman [86] als moderat eingestuft werden. Das 95%-Konfidenzintervall liegt im hinreichenden bis sehr guten Bereich. Der große Bereich des Konfidenzintervalls ist durch die geringe Fallzahl ($n = 32$) zu erklären. Die Spezifität steigt gegenüber dem ersten Datensatz auf 100 %. In der Sensitivität ist eine leichte Verbesserung auf 57,9 % zu erkennen.

Die in Tabelle 8.1 (s. Anhang) dargestellte Kontingenztafel zeigt den zugehörigen Vergleich zwischen PR und Histologie für Readers 2 unter Verwendung von Datensatz 1. Bei Reader 2 liegt in 66 von 116 Fällen (56,9 %) eine übereinstimmende Beurteilung zwischen PR und Histologie vor. Die Inter-Methoden-Reliabilität auf Basis von Cohens Kappa befindet sich mit $K2 = 0,17$ auf ähnlichem Niveau wie bei Reader 1 und muss gemäß Altman [86] als knapp nicht ausreichend eingestuft werden. Das 95%-Konfidenzintervall von $K2$ liegt im nicht

ausreichend bis hinreichenden Bereich. Sowohl Sensitivität (52,5 %) als auch Spezifität (86,7 %) sind vergleichbar zu den Ergebnissen von Reader 1.

Zur Verifikation der Ergebnisse ist die unter Verwendung von Datensatz 2 erzeugte Kontingenztafel in Tabelle 8.2 im Anhang dargestellt. Auch für Reader 2 zeigt sich ein gesteigerter Anteil der übereinstimmenden Beurteilung bezüglich R0/R1 in nun 26 von 32 Fällen (81,3 %). Entsprechend steigt auch Cohens Kappa auf $K_2 = 0,64$) und kann damit gemäß Altman [86] als (knapp) gut eingestuft werden. Das 95%-Konfidenzintervall liegt im hinreichend bis sehr guten Bereich, wobei der weite Bereich des Intervalls durch die geringe Fallzahl von $n = 32$ erklärt werden kann. Analog zu den Ergebnissen der Verifikation für Reader 1, steigt für Reader 2 durch Verwendung von Datensatz 2 die Sensitivität auf nun 68,4 % und die Spezifität ebenfalls auf 100 %.

Um herauszufinden, ob die Übereinstimmung der Diagnose aus PR und Histologie von der Herdbefundgröße abhängt, wurden die Fälle entsprechend der histologisch festgestellten maximalen Herdbefundgröße in Intervallen von 5 mm geclustert und die relative Übereinstimmung (Anteil an Fällen, in denen R0/R1 im Vergleich zur Histologie richtig erkannt wurden) ausgewertet. Die detaillierten Ergebnisse sind Anhang 8.8 zu entnehmen. Eine signifikante Abhängigkeit der Güte der R0/R1-Diagnose der PR von der Herdbefundgröße kann nicht festgestellt werden.

Vergleich DBT und Histologie

Die Kontingenztafel in Tabelle 3.22 zeigt den Vergleich der R0/R1-Diagnose des Reader 1 mit dem Verfahren der DBT zur Histologie. Die Diagnose stimmt in 61 von 107 Fällen zwischen der Histologie und den Ergebnissen aus der DBT von Reader 1 überein (57 %). Die Intermethoden-Reliabilität auf Basis von Cohens Kappa ($K_2 = 0,146$) muss gemäß Altman [86] als nicht ausreichend eingestuft werden. Das 95%-Konfidenzintervall liegt im nicht ausreichend bis hinreichenden Bereich. Die Sensitivität (53,7 %) und Spezifität (83,3 %) der DBT liegen leicht unter denen der PR für Reader 1.

Tabelle 3.22: Kontingenztabelle R0/R1-Vergleich DBT und Histologie (Reader 1)*Kontingenztabelle intraoperative DBT vs Histologie Reader 1 R0/R1 (abs.)*

		DBT Reader 1			Rand- häufigkeit	Spezifität	Sensitivität
		R0	R1				
Histo	R0	10	2	12	0,833		
	R1	44	51	95	0,537		
Rand- häufigkeit		54	53	107			

Kontingenztabelle intraoperative DBT vs Histologie Reader 1 R0/R1 (rel.)

		DBT Reader 1			Rand- häufigkeit
		R0	R1		
Histo	R0	0,093458	0,018692	0,11215	
	R1	0,41121	0,47664	0,88785	
Rand- häufigkeit		0,50467	0,49533		

p0 0,57009

pe 0,49638

K2 0,14638

unterer Wert	-0,03988	oberer Wert	0,33263
--------------	----------	-------------	---------

Die Verifikation der Ergebnisse erfolgt wiederum unter Nutzung von Datensatz 2. Die resultierende Kontingenztabelle ist in Tabelle 3.23 dargestellt. Es zeigt sich ein leicht gesteigener Anteil der übereinstimmenden Beurteilung bezüglich R0/R1 in nun 18 von 29 Fällen (62,1 %). Cohens Kappa steigt entsprechend auf $K2 = 0,334$, muss jedoch gemäß Altman [86] als hinreichend eingestuft werden. Die Spezifität steigt auf 100 % wohingegen die Sensitivität auf 42,1 % abfällt. Die Auswertung für Reader 2 auf Basis von Datensatz 1 ist in der Kontingenztabelle in Tabelle 8.3 im Anhang dargestellt. Die Auswertung zeigt in 66 von 107 bearbeiteten Fällen eine übereinstimmende Beurteilung in Histologie und DBT. Die Inter-Methoden-Reliabilität auf Basis von Cohens Kappa ($K2 = 0,18$) muss gemäß Altman [86] als nicht ausreichend eingestuft werden. Das 95%-Konfidenzintervall liegt im nicht ausreichend bis hinreichenden Bereich.

Tabelle 3.23: Kontingenztabelle R0/R1-Vergleich DBT und Histologie (Reader 1, Datensatz 2)

Kontingenztabelle intraoperative DBT vs Histologie Reader 1 R0/R1 (abs.)

		DBT Reader 1		Rand- häufigkeit		
		R0	R1			
Histo	R0	10	0	10	Spezifität	1,000
	R1	11	8	19	Sensitivität	0,421
Rand- häufigkeit		21	8	29		

Kontingenztabelle intraoperative DBT vs Histologie Reader 1 R0/R1 (rel.)

		DBT Reader 1		Rand- häufigkeit		
		R0	R1			
Histo	R0	0,34483	0	0,34483		
	R1	0,37931	0,27586	0,65517		
Rand- häufigkeit		0,72414	0,27586			

p0 0,62069
pe 0,43044
K2 0,33403
unterer Wert 0,02397 **oberer Wert** 0,64409

Die Sensitivität (58,9 %) ist im Vergleich zu den Ergebnissen von Reader 2 in der PR leicht höher, die Spezifität (83,3 %) ist im Vergleich jedoch ein wenig geringer. Das Resultat der Verifikation mittels Datensatz 2 ist in der Kontingenztabelle in Tabelle 8.4 (s. Anhang) abgebildet. Wie zuvor bei Reader 1 zeigt sich unter Verwendung von Datensatz 2 auch hier ein steigender Anteil der übereinstimmenden Beurteilung bezüglich R0/R1 (22 von 29 Fällen, 75,9 %). Cohens Kappa steigt ebenfalls auf $K2 = 0,542$ und kann damit als moderat eingestuft werden. Die Sensitivität bleibt gegenüber Datensatz 1 mit 63,2 % auf einem vergleichbaren Niveau. Die Spezifität steigt auf 100 %. Analog zur PR kann auch für die DBT keine signifikante Abhängigkeit der Güte der R0/R1-Diagnose von der Herdbefundgröße festgestellt werden. Die detaillierten Ergebnisse sind Anhang 8.8 zu entnehmen.

Tabelle 3.24 fasst die Ergebnisse der Sensitivitäten und Spezifitäten aus dem Vergleich der radiologischen Verfahren mit der Histologie zusammen.

Tabelle 3.24: Vergleich der Sensitivität und Spezifität zwischen den radiologischen Verfahren und der Histologie

	Datensatz 1				Datensatz 2			
	Reader 1		Reader 2		Reader 1		Reader 2	
	SEN	SPE	SEN	SPE	SEN	SPE	SEN	SPE
PR vs. Histo	55,4 %	86,7 %	52,5 %	86,7 %	57,9 %	100 %	68,4 %	100 %
DBT vs. Histo	53,7 %	83,3 %	58,9 %	83,3 %	42,1 %	100 %	63,2 %	100 %

Es ist ersichtlich, dass die Sensitivität und Spezifität für beide radiologischen Verfahren auf einem vergleichbaren Niveau liegen. Den unter Verwendung von Datensatz 1 ersichtlichen, minimalen Vorteilen der DBT bei Reader 2 steht die etwas schlechtere Sensitivität bei Reader 1 entgegen. Bei der Verwendung von Datensatz 2 im Rahmen der Verifikation der Ergebnisse ergeben sich durchweg 100 % Spezifitäten, wohingegen sich die Veränderung der Sensitivität nicht eindeutig darstellt. Es kann daher geschlussfolgert werden, dass das moderate Niveau der Sensitivität in Datensatz 1 nicht auf systematische Fehler (z.B. aufgrund inkorrektur Beurteilung bei unklarer Sachlage) zurückzuführen ist, sondern Methoden-immanent ist.

3.6.4 Sensitivität des R0/R1-Vergleichs auf Randabstandskriterien

Im Kontext der Relevanz für die klinische Praxis ist die Frage bedeutsam, bei welchen Randabstandskriterien die Übereinstimmung zwischen der R0/R1-Diagnose aus den radiographischen Verfahren und der Histologie am größten ist. Auf Basis des vorliegenden Studienmaterials wird in diesem Abschnitt mittels Sensitivitätsanalyse untersucht, bei welchem Randabstandskriterium die beste Übereinstimmung in Bezug auf Sensitivität und Spezifität vorliegt.

Abbildung 3.20 stellt die Sensitivitätsanalyse für die PR und Reader 1 dar.

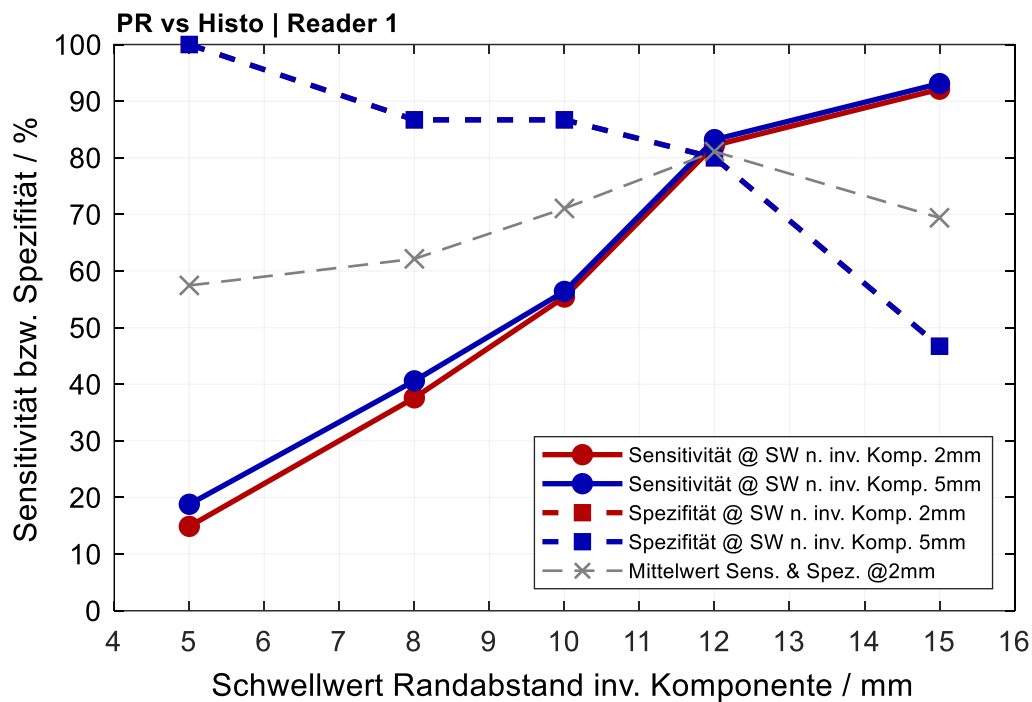


Abbildung 3.20: Sensitivitätsanalyse der R0/R1-Diagnose der PR auf Randabstandskriterien (Reader 1)

Hierbei wurde – ceteris paribus – auf Basis der gemessenen Randabstände die R0/R1-Diagnose des radiologischen Verfahrens für verschiedene Randabstandskriterien neu ausgewertet und die Sensitivität sowie Spezifität über dem R0/R1-Schwellwert (Randabstandskriterium) für die invasive Komponente aufgetragen. Zur Untersuchung des Einflusses des Randabstandskriteriums für die nicht invasive Komponente, wurde ein zweiter Linienzug mit 5 mm Schwellwert ergänzt. Die Randabstandskriterien der Histologie bleiben jeweils unverändert auf dem Standard für diese Arbeit (2 mm bzw. 10 mm). Die analoge Darstellung für Reader 1 und die DBT ist in Abbildung 3.21 zu finden.

Erwartungsgemäß verhalten sich Sensitivität und Spezifität für beide radiologischen Verfahren gegenläufig. Die Sensitivität, also der Anteil der aus Sicht der Histologie durch das radiologische Verfahren richtig identifizierten R1-Fälle, steigt hin zu größeren Schwellwerten für die invasive Komponente, da bei höheren Schwellwerten radiologisch mehr Fälle als R1 klassifiziert werden. Gleiches gilt vice versa für die Spezifität.

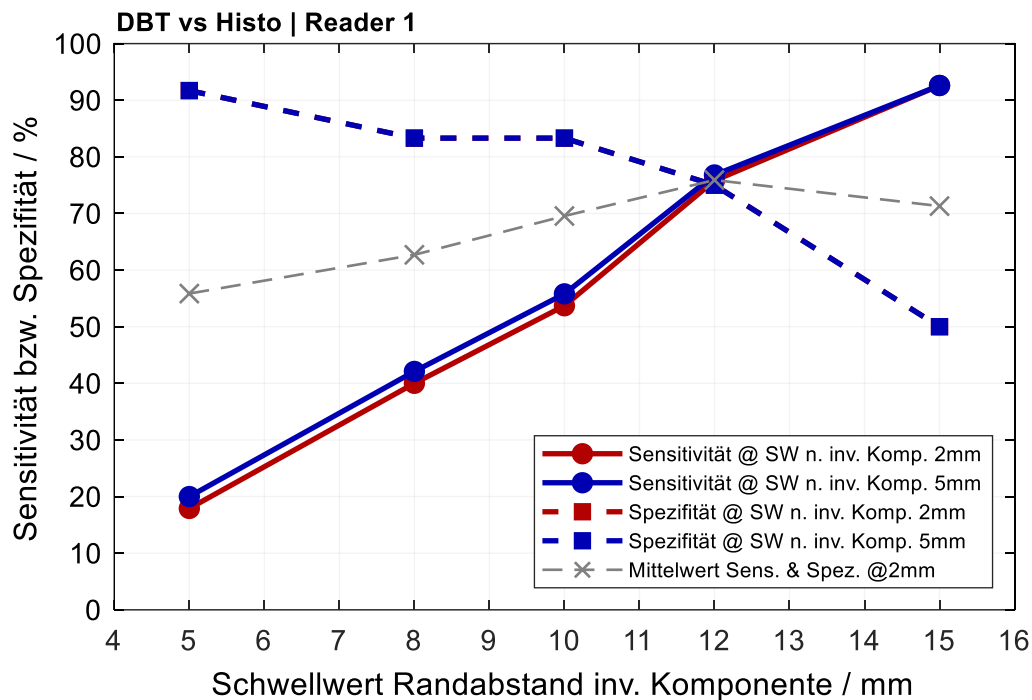


Abbildung 3.21: Sensitivitätsanalyse der R0/R1-Diagnose der DBT auf Randabstandskriterien (Reader 1)

Die Variation des Schwellwertes für die nicht invasive Komponente hat für das in dieser Arbeit vorliegende Studienmaterial eine wesentliche Auswirkung. Dieser Schwellwert wird daher im Folgenden nicht weiter betrachtet. Unter der Annahme gleicher Wichtigkeit von Sensitivität und Spezifität kann der Mittelwert aus beiden Größen herangezogen werden, um das Optimum des Schwellwertes für die invasive Komponente zu beurteilen. Der Schwellwert, bei dem das Maximum der Kurve liegt, stellt das Pareto-Optimum dar. Er ist also der Wert, bei dem eine Verbesserung der einen Zielgröße nicht ohne Verschlechterung der anderen Zielgröße möglich ist. Entgegen der Erwartung liegt das Maximum des Mittelwertes nicht bei 10 mm, sondern bei 12 mm, vgl. Abbildung 3.20. Die Sensitivitätsanalysen für Reader 2 zeigen einen qualitativ ähnlichen Verlauf und können dem Anhang 8.7.4 entnommen werden. Es kann festgestellt werden, dass auf Basis des vorliegenden Studienmaterials ein Versatz von ca. 2 mm in den verwendeten Randabstandskriterien zu einer verbesserten Übereinstimmung führt.

Die Methodik der Sensitivitätsanalysen kann ebenfalls angewendet werden, um den in dieser Arbeit verfolgten Ansatz eines Vergleiches von radiologischen Verfahren und Histologie unter gleichen Randbedingungen zu verifizieren.

Hierzu wurden bei den in Anhang 8.7.5 befindlichen Abbildungen anstelle der angepassten, die originale R0/R1-Diagnose der Histologie verwendet.

Im Quervergleich der Ansätze kann der Mittelwert aus Sensitivität und Spezifität als Gütekriterium verwendet werden. Es ist ersichtlich, dass die Mittelwerte aus Sensitivität und Spezifität für beide radiologischen Verfahren im Vergleich zu der angepassten R0/R1-Diagnose auf geringerem Niveau liegen. Der im Rahmen dieser Arbeit verwendete Ansatz der angepassten R0/R1-Diagnose kann daher auf Basis des höheren Übereinstimmungsniveaus als methodisch richtig bestätigt werden.

Aus der Betrachtung der Lage des Maximums der Mittelwertkurve (Sensitivität und Spezifität) kann jedoch auch geschlussfolgert werden, dass die höchste Übereinstimmung zwischen der originalen R0/R1-Diagnose und den radiologischen Verfahren bei Schwellwerten von ca. 10 bis 12 mm liegt. Die in der radiologischen Praxis gebräuchlichen Schwellwerte können daher im Grunde bestätigt werden.

3.6.5 *Vergleich der Identifikation von Mikrokalk*

In diesem Abschnitt wird die Güte der radiologischen Methoden zur Identifikation von MK beschrieben. Hierzu wurden die Ergebnisse der Fragebögen in Form von Kontingenztafeln aufbereitet und neben dem Übereinstimmungsmerkmal Cohens Kappa die Sensitivität sowie Spezifität gegenüber dem Goldstandard, der Histologie, berechnet.

Vergleich PR und Histologie

Tabelle 3.25 stellt diesen Vergleich für die PR für Reader 1 dar. In 69 von 115 Fällen kommt Reader 1 und die Histologie zum gleichen Ergebnis (60 %). Da diese Übereinstimmung nur wenig oberhalb der zu erwartenden Wahrscheinlichkeit einer Urteilsübereinstimmung bei gegebenen Randhäufigkeiten und bei

stochastisch unabhängigen Beurteilern liegt, ist Cohens Kappa entsprechend niedrig ($K_2 = 0,037$). Die Übereinstimmung muss gemäß Altman [86] als nicht ausreichend eingestuft werden. Die Sensitivität liegt bei 43,8 % die Spezifität bei 62,6 %. Es kann damit das Vorhandensein von MK besser ausgeschlossen als erkannt werden.

Tabelle 3.25: Kontingenztabelle zum Vergleich der Erkennbarkeit von MK in der PR und der Histologie (Reader 1)

Kontingenztabelle intraoperative PR vs Histologie Reader 1 Mikrokalk (abs.)

		PR Reader 1		Rand- häufigkeit		
		nein	ja			
Histo	nein	62	37	99	Spezifität	0,626
	ja	9	7	16	Sensitivität	0,438
Rand- häufigkeit		71	44	115		

Kontingenztabelle intraoperative PR vs Histologie Reader 1 Mikrokalk (rel.)

		PR Reader 1		Rand- häufigkeit
		nein	ja	
Histo	nein	0,53913	0,32174	0,86087
	ja	0,078261	0,06087	0,13913
Rand- häufigkeit		0,61739	0,38261	

p0 0,6

pe 0,58473

K2 0,036781

unterer Wert -0,17883 oberer Wert 0,25239

Für Reader 2 zeigt sich – wie der Kontingenztabelle in Tabelle 8.5 (s. Anhang) zu entnehmen ist – eine mit der Histologie übereinstimmende Beurteilung in 73 von 116 Fällen (62,9 %). Dies liegt auf dem gleichen Niveau wie die Übereinstimmung von Reader 1. Bei der zufallskorrigierten Übereinstimmungsmaßzahl Cohens Kappa ergibt sich ein leicht höherer Wert ($K_2 = 0,141$), wobei dieser gemäß Altman [86] ebenfalls als nicht ausreichend eingestuft werden muss. Die

Sensitivität liegt für Reader 2 bei 62,5 % und damit höher als bei Reader 1. Die Spezifität liegt mit 63 % gleich auf.

Vergleich DBT und Histologie

Tabelle 3.26 stellt den Vergleich der Ergebnisse aus Histologie und DBT für Reader 1 dar. Es kann in 72 von 107 Fällen (67,3 %) eine übereinstimmende Beurteilung zwischen der Histologie und Reader 1 festgestellt werden, was oberhalb des Wertes der PR liegt. Trotzdem liegt Cohens Kappa gemäß Altman [86] nur im nicht ausreichenden Bereich ($K2 = 0,106$), was sich durch das relativ hohe $pe = 0,634$ erklären lässt. Die Sensitivität der DBT liegt mit 43,8 % auf gleichem Niveau wie die Sensitivität der PR für Reader 1.

Tabelle 3.26: Kontingenztabelle zum Vergleich der Erkennbarkeit von MK in der DBT und der Histologie (Reader 1)

Kontingenztabelle intraoperative DBT vs Histologie Reader 1 Mikrokalk (abs.)

		DBT Reader 1		Rand- häufigkeit		
		nein	ja			
Histo	nein	65	26	91	Spezifität	0,714
	ja	9	7	16	Sensitivität	0,438
Rand- häufigkeit		74	33	107		

Kontingenztabelle intraoperative DBT vs Histologie Reader 1 Mikrokalk (rel.)

		DBT Reader 1		Rand- häufigkeit
		nein	ja	
Histo	nein	0,60748	0,24299	0,85047
	ja	0,084112	0,065421	0,14953
Rand- häufigkeit		0,69159	0,30841	

p0 0,6729

pe 0,63429

K2 0,10556

unterer Wert -0,13751

oberer Wert 0,34864

Die Spezifität liegt mit 71,4 % jedoch unterhalb der Spezifität der PR. Tabelle 8.6 bildet die Kontingenztafel für den Vergleich zwischen Histologie und den Ergebnissen der DBT für Reader 2 im Anhang ab. Eine übereinstimmende Beurteilung zwischen Histologie und DBT liegt in 68 von 107 Fällen (63,6 %) vor. Dies liegt auf ähnlichem Niveau wie bei der PR für Reader 2. Die Übereinstimmung der Ergebnisse muss gemäß Altman [86] auf Basis von Cohens Kappa ($K_2 = 0,1$) als nicht ausreichend eingestuft werden. Mit einem Wert von 50 % liegt die Sensitivität der DBT für Reader 2 unterhalb der Sensitivität der PR (62,5 %). Die Spezifität der DBT (65,9 %) liegt jedoch leicht oberhalb der Spezifität der PR (63,0 %).

Die zusammengefassten Ergebnisse für Sensitivität und Spezifität bei der Erkennung von MK sind in Tabelle 3.27 für beide radiologischen Verfahren gegenübergestellt. Hinsichtlich der Fragestellung, ob eines der radiologischen Verfahren besser geeignet ist um MK zu identifizieren, ist festzuhalten, dass die Ergebnisse nicht eindeutig sind. Einer gleichbleibenden Sensitivität beider Verfahren bei Reader 1, steht eine schlechtere Sensitivität der DBT bei Reader 2 entgegen. Bei der Spezifität kehrt sich das Bild zu einem leicht vorteilhafteren Wert für Reader 2 bei der DBT. Diesem steht ein schlechterer Wert für Reader 1 im Vergleich zur PR entgegen. Zudem zeigt die Analyse auf Basis von Cohens Kappa zwischen Histologie und den radiologischen Verfahren eine nicht ausreichende Übereinstimmung.

Tabelle 3.27: Vergleich der Sensitivität und Spezifität von PR und DBT bei der Erkennbarkeit von MK

	Reader 1		Reader 2	
	SEN	SPE	SEN	SPE
PR vs. Histo	43,8 %	86,7 %	62,5 %	63,0 %
DBT vs. Histo	43,8 %	71,4 %	50,0 %	65,9 %

In Abbildung 3.22 ist die Auswertung der subjektiven Einschätzung der Reader hinsichtlich der Erkennbarkeit und Messbarkeit von MK für beide radiologischen Verfahren gegenübergestellt.

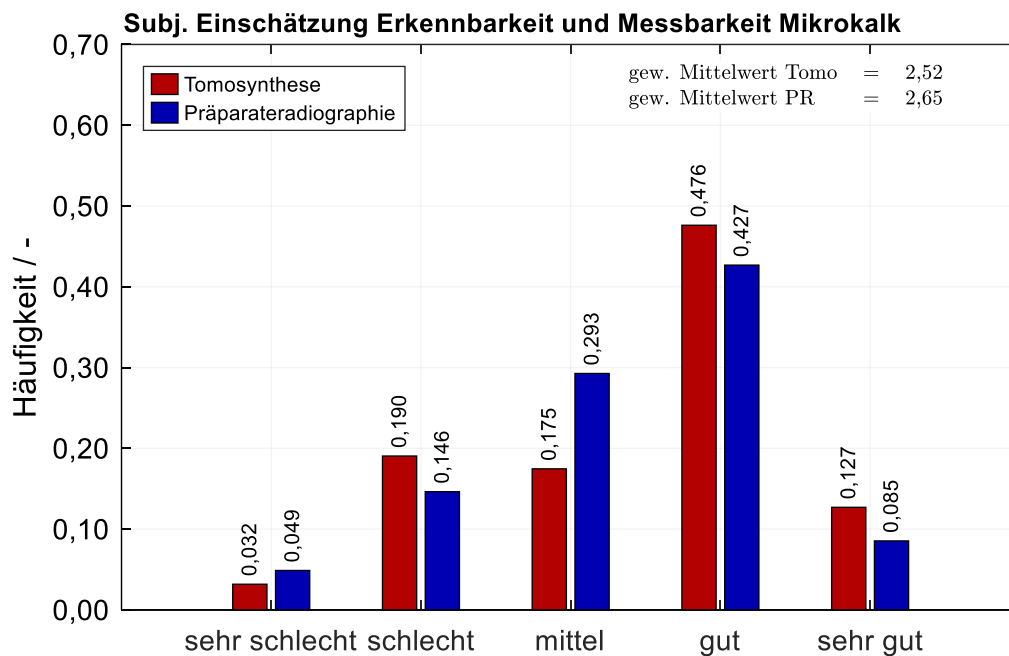


Abbildung 3.22: Subjektive Einschätzung der Reader auf die Frage wie gut MK in DBT und PR erkennbar und messbar sind

Es ist ersichtlich, dass die Erkennbarkeit und Messbarkeit in der DBT häufiger mit ‚gut‘ und ‚sehr gut‘ bewertet wird als in der PR. Allerdings gibt es auch häufiger ‚schlechte‘, dafür aber weniger ‚sehr schlechte‘ und ‚mittlere‘ Bewertungen für die DBT. Aus dem Vergleich der gewichteten Mittelwerte lässt sich ableiten, dass die DBT von den Readern subjektiv bevorzugt wird, obwohl dies anhand der Ergebnisse der Konkordanzanalyse zwischen den Verfahren und der Histologie nicht bestätigt werden kann.

Da sowohl Sensitivität und Spezifität deutlich unter 100 % liegen und die Übereinstimmung zur Histologie nicht ausreichend ist, ist festzuhalten, dass weder PR noch DBT in der Lage sind MK zuverlässig zu identifizieren.

3.6.6 Zeitlicher Aufwand zur Analyse der Präparate in den radiologischen Verfahren

Zur Beurteilung des zeitlichen Aufwands wurde zum einen durch die Reader die Bearbeitungsdauer gestoppt sowie eine subjektive Einschätzung zum Zeitaufwand in fünf Bewertungsstufen von sehr hoch bis sehr gering abgegeben. Die Auswertung der subjektiven Einschätzung für beide radiologischen Verfahren ist in Abbildung 3.23 dargestellt. Subjektiv wird von den Readern die DBT über alle Fälle hinweg als weniger zeitaufwändig eingeschätzt. Dies zeigt sich insbesondere in den vermehrten Bewertungen in der Kategorie ‚gering‘ sowie den weniger häufigen Bewertungen in der Kategorie ‚hoch‘ und spiegelt sich ebenfalls im gewichteten Mittelwert wider. Die Zeitmessung ergibt im Mittel über alle Fälle ähnliche Bearbeitungszeiten von 2:40 min für die DBT und 2:35 min für die PR. Die DBT wird daher bei vergleichbarem Zeitaufwand (3 % zugunsten der PR) von den Readern subjektiv bevorzugt.

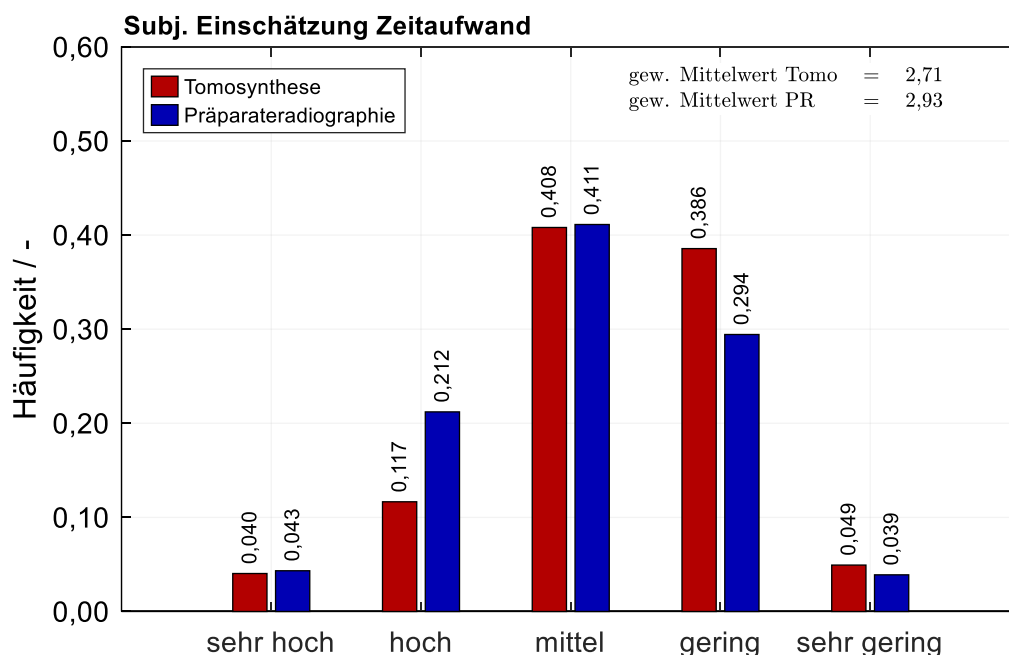


Abbildung 3.23: Subjektive Einschätzung der Reader auf die Frage wie der zeitliche Aufwand der Messdatenerhebungen in DBT und PR empfunden wurde

3.7 Zusammenfassung der Ergebnisse

Im Hinblick auf Alter, Brustdichte, Entität und Lokalisation ist das Studienmaterial dieser Arbeit vergleichbar mit anderen Studien und kann daher als repräsentativ eingestuft werden.

Die Auswertung der präoperativen Mammographie ergibt eine geringe mittlere Abweichung zwischen den Readern bei der Bestimmung der Herdbefundgröße (0,2...1,3 mm), jedoch eine hohe Streuung zwischen den Messwerten. Die LoA liegen im Bereich zwischen -9,6...10,2 mm und verletzen damit das im Rahmen dieser Arbeit aus klinischer Sicht festgelegte Gütekriterium von ± 5 mm. Der Inter-Reader-Vergleich bei der Messung von MK zeigt abweichend hierzu höhere mittlere Abweichungen sowie ebenfalls eine hohe Streuung zwischen den Readern. Dieser kann aufgrund des geringen Stichprobenumfangs infolge weniger erkannter MK nicht abschließend bewertet werden. Die Übereinstimmung zwischen den Readern wird folglich als nicht ausreichend eingestuft.

Im Inter-Reader-Vergleich zur Präparatgröße ergeben sich für das Verfahren der PR geringe mittlere Abweichungen im Bereich von 0,2...1,6 mm. Die LoA liegen zwischen -13,5...16,8 mm und überschreiten das Gütekriterium. Der Vergleich beider Reader bei der Bestimmung der Herdbefundgröße zeigt über alle Messgrößen hinweg eine mittlere Abweichung von 0,9...1,4 mm, die LoA liegen zwischen -12,0...14,6 mm. Bei der Messung des Randabstandes ergeben sich mittlere Abweichungen im Intervall von -2,8...2,3 mm. Die LoA liegen zwischen -23,8...22,4 mm und überschreiten das Gütekriterium. Trotz grundsätzlich guter Korrelation und im Mittel kleiner Abweichungen der Messergebnisse der Reader muss auf Basis der hohen Streuung der Messwerte umeinander festgehalten werden, dass es keine vollständige Übereinstimmung zwischen den Messungen der beiden Reader gibt. Folgerichtig liegt die Inter-Reader-Reliabilität bei der R0/R1-Diagnose für die PR im moderaten Bereich ($K2 = 0,57$).

Für das Verfahren der DBT ergeben sich im Inter-Reader-Vergleich zur Messung der Präparatgröße ebenfalls geringe mittlere Abweichungen im Bereich von 1,3...1,4 mm. Die LoA liegen in einem Bereich von -8,7...11,2 mm und damit niedriger als in der PR. Dennoch überschreitet der Bereich das Gütekriterium. Bei der Bestimmung der Herdbefundgröße zeigt der Vergleich beider Reader

über alle Messgrößen hinweg eine mittlere Abweichung von 1,3...1,8 mm, die LoA liegen zwischen -7,3...10,8 mm. Die Streuung zwischen den Readern liegt damit auch bei der Herdbefundgröße unterhalb der Streuung der PR. Die mittleren Abweichungen zwischen den Messungen des Randabstandes liegen im Intervall von -1,6...0,1 mm. Die LoA liegen zwischen -19,4...18,7 mm. Bei der Bestimmung des Randabstandes mittels DBT liegt die Inter-Reader-Streuung auf dem Niveau der Streuung der PR. Auch hier wird das Gütekriterium überschritten. Auf Basis der hohen Streuung der Messwerte umeinander muss festgehalten werden, dass es keine vollständige Übereinstimmung zwischen der Messung der beiden Reader gibt. Trotzdem liegt die Inter-Reader-Reliabilität für die DBT bei der R0/R1-Diagnose im guten Bereich ($K2 = 0,72$) und damit deutlich höher als bei der PR.

Im direkten Inter-Methoden Vergleich der aus PR und DBT abgeleiteten R0/R1-Diagnose ergeben sich für beide Reader nur eine moderate bis hinreichende Übereinstimmung (Reader 1 $K2 = 0,5$, Reader 2 $K2 = 0,39$). Hieraus kann zum einen abgeleitet werden, dass es einen belegbaren Unterschied in der diagnostischen Güte zwischen PR und DBT gibt, zum anderen aber auch, dass der Reader einen Einfluss hierauf hat.

Der Vergleich beider radiologischer Verfahren mit dem Goldstandard Histologie ergibt mit Blick auf die Messung der maximalen Herdbefundgröße kleinere mittlere Abweichungen (DBT: 0,5/2,4 mm, PR: 1,6/2,7 mm jeweils für Reader 1 / Reader 2), sowie engere LoA (DBT: -9,5...10,4 / -8,6...13,4 mm, PR: -12,1...15,2 / -8,7...14,0 mm jeweils für Reader 1 / Reader 2) zugunsten der DBT für beide Reader. Es kann daher geschlossen werden, dass die DBT relativ zur PR besser geeignet ist, die Herdbefundgröße zu bestimmen. Trotzdem verletzt auch die DBT das im Rahmen dieser Arbeit aus klinischer Sicht festgelegte Gütekriterium von ± 5 mm für die LoA.

Auf Basis der beschriebenen Limitationen der Randabstandbestimmung in der Histologie ist der Vergleich der Randabstände zwischen dem Goldstandard und den radiologischen Verfahren nur eingeschränkt möglich. Unter Nutzung der im Rahmen dieser Arbeit vorliegenden Messwertpaaren für Randabstände < 10 mm

zeigt sich das Bild einer nicht ausreichenden Korrelation und einer hohen Streuung, sodass eine robuste R0/R1-Diagnose nicht möglich erscheint.

Im Vergleich der aus den radiologischen Verfahren abgeleiteten R0/R1-Diagnose mit der Histologie ergibt sich unter Verwendung einheitlicher Randabstandskriterien eine Sensitivität von 55,4 % / 52,5 % für die PR und von 53,7 % / 58,9 % für die DBT (jeweils für Reader 1 / Reader 2). Die Spezifität liegt für beide Reader für die PR bei 86,7 % und bei 83,3 % für die DBT. Damit liegen beide Verfahren auf einem vergleichbaren Niveau, das in Bezug auf die Sensitivität moderat ausfällt. Die Verwendung eines auf eindeutige Fälle eingeschränkten Datensatzes bestätigt das moderate Niveau und schließt subjektive Einflüsse durch Interpretationsspielraum bei der Beurteilung der Fälle aus. Eine signifikante Abhängigkeit der Güte der R0/R1-Diagnose von der Herdbefundgröße kann für beide radiologische Verfahren nicht festgestellt werden. Im Rahmen einer Sensitivitätsanalyse konnte eine weitere Verbesserung der Güte der R0/R1-Diagnose der radiologischen Verfahren bei Verwendung eines Versatzes von ca. +2 mm in den Randabstandskriterien gegenüber der Histologie identifiziert werden.

Auf Basis des Vergleichs zur Erkennbarkeit und Messbarkeit der Größe von MK kann kein eindeutiges Ergebnis abgeleitet werden. Die PR weist eine leicht höhere Sensitivität (PR: 43,8/62,5 %, DBT: 43,8 / 50,0 % jeweils für Reader 1 / Reader 2) und eine leicht höhere Spezifität (PR: 86,7 / 63,0 %, DBT: 71,4 / 65,9 % jeweils für Reader 1 / Reader 2) als die DBT auf. Bei Betrachtung der Übereinstimmung der Ergebnisse mit der Histologie muss mit Blick auf das zufallskorrigierte Übereinstimmungsmaß Cohens Kappa jedoch eine nicht ausreichende Übereinstimmung festgehalten werden. Zudem wird die DBT in der subjektiven Einschätzung der Reader gegenüber der PR knapp bevorzugt.

Die DBT wird von den Readern neben dem Kriterium der Erkennbar,- bzw. Messbarkeit von MK auch mit Blick auf den zeitlichen Aufwand sowie auf die Messbarkeit des Herdbefundes präferiert.

4 Diskussion

4.1 Vergleich zum Stand der Wissenschaften

Im nachfolgenden Abschnitt werden die Ergebnisse der vorliegenden Arbeit mit verfügbaren Ergebnissen aus der Literatur verglichen und eingeordnet. In der Literatur finden sich nach aktuellem Kenntnisstand nur wenige detaillierte Inter-Reader-Vergleiche für den intraoperativen Einsatz der Präparateradiographie und Präparatetomosynthese, die sich eignen, die im Rahmen dieser Arbeit dargelegten Erkenntnisse zu verifizieren. Entsprechende Inter-Reader-Vergleiche existieren im Wesentlichen im Bereich des präoperativen Einsatzes, also der Darstellung der gesamten Brust [90, 91]. Hierbei wurde neben dem Einsatz zur BI-RADS-Klassifikation auch die Eignung der Verfahren zur Charakterisierung von MK untersucht [92]. Entsprechend fokussiert sich dieser Abschnitt nach einer kurzen Übersicht zum Inter-Reader-Vergleich bei der Randabstandsmessung hauptsächlich auf den Vergleich der Verfahren mit dem Goldstandard, der Histologie.

Inter-Reader-Vergleich Randabstandsmessung

In Romanucci et al. [93] wird der Inter-Reader-Vergleich anhand des minimalen Randabstandes mittels Bland-Altman-Analyse zwischen zwei Readern durchgeführt. Dieser basiert auf einem Stichprobenumfang von $n = 170$. Abweichend zu der vorliegenden Arbeit wurden in Romanucci et al. die Randabstände nicht in den Raumrichtungen orientiert gemessen. Der jeweils kleinste Randabstand in der CC- und der LL-Ebene wurde von den Readern identifiziert, eine Aussage über den gleichen Messort oder die Messrichtung wird nicht getroffen. Für die PR ergeben sich in Romanucci et al. mittlere Abweichungen von $-0,07$ mm in der CC-Ebene und $0,20$ mm in der LL-Ebene. Die Standardabweichungen liegen bei $5,99$ mm und $6,42$ mm. Für die DBT werden $-0,07$ mm in der CC-Ebene und $-0,12$ mm in der LL-Ebene ausgewiesen, wobei die Standardabweichungen bei $6,13$ mm und $6,54$ mm liegen. Im Rahmen der vorliegenden Arbeit wurden die AP- und die LL-Ebene ausgewertet. Es ergeben sich über die verschiedenen

Randabstandsmessgrößen hinweg in der LL-Ebene eine mittlere Abweichung im Bereich von -2,83...0,52 mm und zugehörige Standardabweichungen zwischen 4,66...11,15 mm. Bei der DBT ergeben sich in der LL-Ebene eine mittlere Abweichung von -1,57...-0,34 mm sowie eine Standardabweichung zwischen 4,26...8,97 mm. Die im Rahmen der vorliegenden Arbeit ermittelte Übereinstimmung liegt damit zwar in einer ähnlichen Größenordnung, ist jedoch für einzelne Messwerte größer und mitunter mit einer höheren Streuung behaftet. Es sei jedoch angemerkt, dass sich auch bei der nach Einschätzung von Romanucci et al. „exzellenten“ Inter-Reader-Übereinstimmung am Beispiel der LL-Ebene für die PR LoA von -12,4...12,8 mm und für die DBT LoA von -12,9...12,7 mm ergeben. Diese erscheinen mit Blick auf einen im histologischen Mittel 11 mm großen Randabstand jedoch recht hoch und verletzen das im Rahmen dieser Arbeit festgelegte klinisch relevante Gütekriterium von ± 5 mm. Eine robuste Einschätzung der R0/R1-Diagnose -im Sinne eines gleichen Ergebnisses für unterschiedliche Reader- ist auf Basis der im Rahmen dieser Arbeit und der Arbeit von Romanucci et al. ermittelten Streuungen bei der Randabstandsmessung trotz im Mittel guter Übereinstimmungen fraglich.

Messung der Herdbefundgröße

In Seo et al. [94] wird die gemessene maximale Herdbefundgröße von drei Readern für die PR und DBT mit der Größe des Herdbefundes aus der Histologie verglichen. Für die DBT wird eine mittlere Abweichung von 1,6 mm bei einer Standardabweichung von 14,5 mm festgestellt. In der PR ergibt sich eine größere mittlere Abweichung von 4,4 mm sowie eine höhere Standardabweichung von 15,4 mm. Die Ergebnisse dieser Arbeit auf Basis der maximalen Herdbefundgröße bestätigen die Erkenntnis, dass die DBT sowohl eine geringere mittlere Abweichung (DBT: 0,5/2,4 mm, PR: 1,6/2,7 mm jeweils für Reader 1 / Reader 2) als auch eine kleinere Streuung (DBT: 5,1/5,6 mm, PR: 7,0/5,8 mm jeweils für Reader 1 / Reader 2) im Vergleich zur PR aufweist. Ebenfalls kann die in [94] festgestellte Überschätzung der Herdbefundgröße in beiden Verfahren bestätigt werden. Dennoch erscheint die Streuung in Seo et al. gegenüber der

vorliegenden Studie deutlich erhöht. Eine kritische Diskussion bezüglich der klinischen Relevanz der Streuung ist [94] jedoch nicht zu entnehmen. Die aus den Standardabweichungen zu ermittelnden LoA für die Studie von Seo et al. sind entsprechend hoch (PR: -25,8...34,6 mm, DBT: -26,8...30,0 mm) und verletzen die im Rahmen dieser Arbeit festgelegten Kriterien deutlich, so dass auch diese Erkenntnis als bestätigt angesehen werden kann.

Messung von Randabständen

In Romanucci et al. [93] ist ein Vergleich der minimalen Randabstandsmessungen aus PR und DBT mit den histologischen Befunden dargestellt. In der LL-Ebene ergeben sich eine mittlere Abweichung zur Histologie von 1,56 mm für die PR und 0,25 mm für die DBT. Die Standardabweichungen liegen bei 4,13 mm (PR) sowie 2,21 mm (DBT).

Im Rahmen dieser Arbeit wurden die Randabstände nach den Raumrichtungen orientiert gemessen und verglichen. Hierbei liegen die mittleren Abweichungen in der LL-Ebene für die PR über beide Reader zwischen -3,18...1,68 mm und die Standardabweichungen zwischen 2,73...4,43 mm. Für die DBT ergeben sich mittlere Abweichungen von -2,58...1,65 mm sowie Standardabweichungen zwischen 1,96...4,6 mm. Die Werte sind damit zwar in der gleichen Größenordnung jedoch etwas höher als in [93]. Sie bestätigen die Aussage aus Romanucci et al. im Hinblick auf einen substanziellen Unterschied zwischen DBT und PR nicht. Eine kritische Diskussion der klinischen Relevanz der Streuung findet in [93] nicht statt. Aus den Ergebnissen von Romanucci et al. sind LoA zwischen -6,5...9,7 mm für die PR und -4,1...4,6 mm für die DBT. Damit erfüllt die DBT das im Rahmen dieser Arbeit aus klinischer Sicht festgelegte Gütekriterium, die PR jedoch nicht.

Amer et al. stellen in [83, 95] ebenfalls eine Übereinstimmungsanalyse zwischen der Randabstandsmessung aus der Histologie und den beiden radiologischen Verfahren PR und DBT vor. Hierbei werden der minimale Randabstand sowie dessen Orientierung von zwei Readern bestimmt. Mit Blick auf die Orientierung des kleinsten Randabstandes gelingt es den Readern mittels PR in 44 % (Reader

1) bzw. 36 % (Reader 2) und mittels DBT in 68 % (Reader 1) bzw. 69 % (Reader 2), diese korrekt zu bestimmen. Die Aussagen zur Sensitivität und Spezifität der Verfahren werden im Anschluss auf Basis der hinsichtlich Orientierung des minimalen Randabstandes korrekt eingeordneten Fälle vorgenommen (vgl. nächsten Abschnitt). Für die Übereinstimmungsanalyse bei dem Betrag des minimalen Randabstandes wird diese Differenzierung nicht erwähnt. Für die PR werden eine mittlere Abweichung von 2,8/3,5 mm sowie eine Standardabweichung von 2,8/3,7 mm (jeweils Reader 1 / Reader 2) angegeben. Die mittlere Abweichung bei der DBT beträgt hingegen 0,9/0,6 mm bei einer Standardabweichung von 2,8/2,8 mm (jeweils Reader 1 / Reader 2). Die Zahlenwerte bestätigen die Größenordnung der im Rahmen dieser Arbeit ermittelten Werte, wenngleich im Rahmen dieser Arbeit nicht dieselbe vergleichende Aussage zur Überlegenheit der DBT mit Blick auf die Randabstandsmessung getroffen werden kann. Die sich aus der Arbeit von Amer et al. ergebenden LoA verletzen jedoch die im Rahmen dieser Arbeit festgelegten Gütekriterien. Im Rahmen der Arbeit von Amer et al. wird die Streubreite mit Blick auf die klinische Relevanz nicht weiter diskutiert. Vor dem Hintergrund der Ergebnisse dieser Arbeit sowie der verfügbaren Literatur muss festgehalten werden, dass die Streuung der radiologischen Randabstandsmessverfahren mitunter zu hoch ist, um eine robuste R0/R1-Diagnose stellen zu können. Wie die Ergebnisse von [93] zeigen, gelingt es jedoch unter bestimmten Umständen auch, die Gütekriterien einzuhalten.

Ableitung von R0/R1-Diagnose

Aufgrund der Möglichkeit zur Ableitung einer Aussage über die Vollständigkeit des exzidierten Tumors, bevor das Operationsgebiet verschlossen wird, ist die intraoperative PR ein wichtiger diagnostischer Bestandteil der brusterhaltenden Operation. Um mögliche Re-Operationen, mit den daraus resultierenden wirtschaftlichen und individuellen Konsequenzen für die Patientin zu vermeiden, ist eine hohe Sensitivität der PR erforderlich. Verschiedene Studien zeigen, dass die Sensitivität der PR in einem weiten Bereich zwischen ca. 25 % und 75 % liegt [70, 75, 83, 95–97] und unter anderem von den festgelegten Randabstands-

kriterien [96] abhängt. Durch Einschränkungen im Stichprobenumfang, z.B. nur durch Auswertung derjenigen Fälle bei denen die Radiologin den kleinsten Randabstand in der richtigen Messebene identifiziert, kann die ausgewiesene Sensitivität vergrößert werden [83, 95, 98]. Die moderaten Werte der Sensitivität und deren große Spannbreite über die veröffentlichten Studien zeigen, dass das Verfahren der interoperativen PR zwar eine Verbesserung der Re-Exzisionsrate herbeiführen kann [98], jedoch nicht ausreichend genau ist, um eine robuste Vorhersage zu der korrekten R0/R1-Diagnose bezogen auf die spätere Histologie treffen zu können. Mitunter wird sogar infrage gestellt, ob vor dem Hintergrund der durch die radiologische Beurteilung verlängerten Operationsdauer und der entstehenden zusätzlichen Kosten überhaupt ein Vorteil für die Patientinnen entsteht [99].

Die DBT kann als zusätzliche präoperative Bildgebung die Sensitivität und Spezifität der Mammographie erhöhen [8]. Es gilt daher die Hypothese zu prüfen, ob die DBT bei der intraoperativen Beurteilung des Präparats eine Verbesserung in Sensitivität und Spezifität gegenüber der PR herbeiführen kann.

In den Studien von Schultz-Wendtland et. al. zeigt sich eine statistisch signifikante Erhöhung der Sensitivität durch Einsatz der DBT (bis zu 8 %) im Vergleich zur PR [53, 55]. Die Auswertung der ohne Kompression erzeugten Bilder von 100 Präparaten in [53] wurde von 3 Radiologinnen unter Verwendung des Randabstandskriteriums von 10 mm durchgeführt. Hierbei ergeben sich Sensitivitäten von 78,3 % für die PR und von 86,0 % für die DBT. Die Spezifitäten liegen jeweils bei 100 %.

Eine zusätzliche Vergrößerung der PR-Bilder von 1:1 auf 1:1,8 erhöht die Sensitivität über die Reader gemittelt von 78,3 % auf 83 %.

In der Arbeit von Amer [83], die auszugsweise auch in [95] veröffentlicht wurde, werden 102 Patientinnen mit nicht-palpablem Brustkrebs brusterhaltend operiert und die Präparate mittels PR und DBT von zwei Radiologinnen begutachtet. Für die Beurteilung des R0/R1-Status wird ein Sicherheitsabstand für DCIS von ≥ 5 mm und für die invasive Komponente von ≥ 1 mm verwendet. Die Aufnahme des Bildmaterials erfolgt ohne Kompression. In der PR ergibt sich eine Sensitivität von 23,2 % und eine Spezifität von 43,9 %. Für die DBT liegen diese Werte

bei 53,6 % (Sensitivität) und 50 % (Spezifität). Aufgrund der Feststellung, dass die mit der Histologie übereinstimmende Orientierung des kleinsten Randabstandes für die PR in 44 % / 36 % und für die DBT in 68 %/69 % (jeweils Reader 1 / Reader 2) der Fälle korrekt identifiziert werden konnte, wird die Übereinstimmung der R0/R1-Diagnose auch auf Basis der hinsichtlich Orientierung korrekt eingestuften Fälle ausgewiesen. Unter Verwendung der entsprechend kleineren Stichprobe ergibt sich eine durchschnittliche Sensitivität von 77 % für die DBT und von 62 % für die PR. Die Spezifitäten ergeben sich zu 97 % für die PR und zu 77 % für die DBT.

Im Rahmen der vorliegenden Arbeit (Randabstandskriterium ≥ 10 mm (invasive Komponente) und von ≥ 2 mm (DCIS), Aufnahmen mit Kompression) wurde eine über die Reader gemittelte Sensitivität von 54 % für die PR und von 56,3 % für die DBT ermittelt. Die entsprechend aufbereitete Spezifität ergibt sich zu 86,7 % (PR) und 83,3 % (DBT). Die Werte liegen damit oberhalb der unkorrigierten Werte von Amer et al., jedoch unterhalb der Werte von Schultz-Wendtland et al. Ein substanzieller Unterschied zwischen DBT und PR ist damit im Gegensatz zu den anderen Studien nicht festzustellen. Auf Basis der Erkenntnis, dass das gewählte Randabstandskriterium einen Einfluss auf die Sensitivität und Spezifität hat [96] können die Zahlenwerte jedoch nur bedingt direkt miteinander verglichen werden. Zudem ist auffällig, dass sowohl Amer als auch Schultz-Wendtland et al. die radiologischen Aufnahmen ohne Kompression durchgeführt haben. Gemäß Graham et al. [79] bewirkt die Kompression zwar eine verbesserte Auflösung einer nicht-palpablen Läsion, sie hat jedoch auch eine verstärkende Wirkung auf den Höhenverlust (Pancake-Phänomen) des Präparats und damit auch auf die radiologisch gemessenen Abstände.

Im Rahmen dieser Arbeit konnte gezeigt werden, dass ein Randabstandskriterium von 10...12 mm zu einer möglichst hohen Übereinstimmung zwischen histologischem und radiologischem Befund führt und dass ein Versatz in den Randabstandskriterien zwischen radiologischen Verfahren und Histologie die Sensitivität und Spezifität weiter erhöhen kann. Im Hinblick auf die praxisnahe klinische Relevanz sollten diese Feststellungen in zukünftigen Studien näher untersucht werden.

4.2 Kritische Diskussion der Ergebnisse und Einflussfaktoren

Im folgenden Abschnitt werden die Einflussfaktoren auf die in dieser Arbeit gewonnenen Erkenntnisse sowie die möglicherweise resultierenden Abweichungen diskutiert. Dies bildet zugleich den Kern der Limitationen bezüglich der Aussagen der vorliegenden Arbeit. Die Einflussfaktoren lassen sich entsprechend ihres Ursprungs in drei Kategorien, die prozessualen, die biologisch/chemischen und die allgemeinen/statistischen Einflussfaktoren einteilen.

Unter prozessualen Einflussfaktoren werden potenzielle Fehlerquellen verstanden, die sich entlang des verwendeten Vorgehens ergeben können und die durch einen geänderten Ablauf möglicherweise vermieden werden könnten. Entlang des Ablaufes von Operation und Markierung des Präparats über die Erstellung der radiologischen Bilder bis hin zur Histologie ergibt sich zunächst bei der Einspannung des Präparats für die Bildgebung ein erster Einflussfaktor. Es kann im Rahmen dieser Arbeit nicht ausgeschlossen werden, dass zwischen der DBT und PR ein Einfluss auf die Positionierung vorgenommen worden ist. Dies lässt sich anhand der technischen Angaben zur Dicke des Präparats nach Kompression sowie zur verwendeten Kraft für die Kompression ableiten, die mitunter zwischen PR und DBT unterschiedlich sind. Dies ist zwar im Inter-Reader-Vergleich innerhalb einer Methode unerheblich, kann aber im Inter-Methoden-Vergleich dazu führen, dass Abweichungen in den Messwerten auftreten. Beim Vergleich von Messwerten zwischen den radiologischen Verfahren und der Histologie in einer definierten Ebene könnte es trotz der Verwendung der gleichen Faden-/Clipmarkierung zu Abweichungen kommen. Insbesondere im kleinen Messbereich der Randabstände können kleine Winkeldifferenzen in der räumlichen Orientierung bereits große Auswirkungen auf den Messwert haben.

In der subjektiven Auswahl des Messortes (d.h. des Anfangs- und Endpunktes) durch die Reader und die Pathologin liegt ein weiterer prozessualer Einflussfaktor auf die Ergebnislage vor. So kann ein unterschiedlicher Messort beim Inter-Reader-Vergleich zu unterschiedlichen Messwerten führen. Bei der DBT kommt im Vergleich zur PR neben der Auswahl des Messortes in der definierten Ebene des jeweils vorliegenden Bildes noch die größere Anzahl an Ebenen hinzu. Auch in der Histologie wird der Messort von der Pathologin bestimmt und muss nicht

zwingend mit dem der radiologischen Verfahren übereinstimmen. Im Rahmen dieser Arbeit wurden die Messorte nicht dokumentiert. Um hier in zukünftigen Studien eine Aussage treffen zu können, wird empfohlen zumindest für die radiologischen Verfahren die Koordinaten des Messortes digital zu erfassen und gegenüberzustellen. Nach persönlicher Einschätzung der Autorin sind unterschiedlich gewählte Messorte die größte prozessuale Quelle für Abweichungen zwischen den Messergebnissen. Möglicherweise können computerunterstützte Auswerte-Algorithmen helfen, diese Limitation zu umgehen, indem standardisiert und automatisiert eine Vielzahl von Messwerten aufgenommen und anschließend automatisiert der kleinste Randabstand einer Bildebene bestimmt wird.

Als weitere prozessuale Limitation muss im Rahmen dieser Arbeit darauf hingewiesen werden, dass die histologische Befundung von unterschiedlichen Pathologinnen durchgeführt wurde. Obwohl dies dem klinischen Standard entspricht, sind subjektive Einflüsse z.B. bei der Messortauswahl, nicht ausgeschlossen. Bei zukünftigen Studien sollte dies zur Verbesserung der Genauigkeit ausgeschlossen sein bzw. durch mehrfache Beurteilung der Präparate durch unterschiedliche Pathologinnen näher untersucht werden.

Den prozessualen Limitationen überlagert sind biologisch/chemische Veränderungsprozesse, die das Präparat in der Zeit zwischen Operation, radiologischer Untersuchung und Histologie verändern. Insbesondere das von Graham et. al. beschriebene „Pancake“-Phänomen hat erheblichen Einfluss auf die Übereinstimmung der Messwerte [79].

Möglicherweise könnte ein Einspannen des Präparats bereits im Operationssaal in ein vorgefertigtes System hilfreich zur weiteren Prozessoptimierung sein. Hier würde einmalig die korrekte räumliche Zuordnung des Präparats von den Chirurgen definiert und von allen weiteren Fachabteilungen eingespannt empfangen bzw. beurteilt werden. Dies würde neben der primär definierten räumlichen Orientierung eine weitere Manipulation am Gewebe selbst vermeiden. Auch dem von Graham et. al. beschriebenen „Pancake“-Phänomen, dem Volumenverlust insbesondere von gesundem Gewebe, würde ein eingespanntes Präparat weniger stark unterliegen [79]. Des Weiteren könnte eine radiographische Beurteilung des Präparats bereits im Operationssaal, neben der kürzeren Operationszeit mit

daraus resultierenden wirtschaftlichen und individuellen Vorzügen für die Patientin, mögliche Fehler während des Transportes in eine weitere Abteilung minimieren. In Partain et al. wird zudem gezeigt, dass die von den Chirurginnen selbst durchgeführte intraoperative DBT im Operationssaal zu einer Senkung der Rate an Re-Operationen führt [100]. Nachteilig sind hier die erforderlichen Strahlenschutzvorkehrungen im Operationssaal, welche bei der Durchführung in der radiologischen Abteilung entfallen.

Neben den zuvor beschriebenen möglichen prozessualen und biologisch/chemischen Einflussfaktoren bestehen weitere allgemeine und statistische Limitationen. So wurden die Daten der vorliegenden Arbeit an einem monozentrischen Standort retrospektiv ausgewertet. Auch wenn dies dem klinischen Standard entspricht, war eine gezielte Einflussnahme auf die prozessualen Abläufe im Rahmen dieser Arbeit nicht möglich. Darüber hinaus sind die Ergebnisse dieser Arbeit nur mittels radiologischer Geräte eines Herstellers erzielt worden. Eine mögliche Verallgemeinerung ist daher entsprechend von den technischen Eigenschaften anderer Gerätschaften sowie von deren Weiterentwicklung abhängig. Der im Rahmen dieser Arbeit genutzte Stichprobenumfang sowie die Anzahl der Reader könnte in zukünftigen, prospektiven Studien weiter erhöht sowie der prozessuale Ablauf weiter standardisiert werden, um die Ergebnisse auf einer breiteren Basis abzusichern.

Abschließend lässt sich zusammenfassen, dass die vorliegende Studie aus akademischer Perspektive zwar viele Limitationen aufweist, jedoch unter Einhaltung von klinischen Standards entstanden ist. Damit können die Ergebnisse der Arbeit als realitätsnahes Abbild interpretiert werden.

5 Zusammenfassung und Ausblick

Auf Basis der Ergebnisse und der kritischen Diskussion der Ergebnisse, werden in diesem Abschnitt die Schlussfolgerungen mit Bezug auf die zugrundeliegende Ziel- und Fragestellung zusammengefasst sowie ein Ausblick auf künftige Forschungsfelder formuliert.

Mit Blick auf die Evaluation der intraoperativen diagnostischen Wertigkeit der DBT (3D-Radiographie) im Vergleich zur 2D-Präparateradiographie kann festgehalten werden, dass die DBT eine höhere Inter-Reader-Reliabilität als die PR für die Messgrößen selbst sowie auch für die daraus abgeleiteten R0/R1-Diagnose aufweist. MK können in beiden Verfahren nicht ausreichend genau vermessen werden. Im Vergleich zum Goldstandard, der Histologie, liegen die ermittelten Sensitivitäten und Spezifitäten beider Verfahren im Streuband der Literaturergebnisse. Im Gegensatz zu anderen Forschungsergebnissen kann jedoch kein substantieller Vorteil für die DBT abgeleitet werden. Im Allgemeinen kann daher festgehalten werden, dass die Ergebnisse beider Verfahren deutlich von unterschiedlichen Randbedingungen wie der Beschaffenheit des Studienmaterials, den Readern, dem prozessualen und zeitlichen Ablauf der radiologischen und histologischen Befundung sowie den gewählten Randabstandskriterien und der Kompression abhängig sind. Zur Ableitung von validen Aussagen müssen – wie in dieser Arbeit gezeigt – einheitliche Randabstandskriterien zwischen radiologischen Verfahren und Histologie verwendet werden. Die im Rahmen dieser Arbeit aufgezeigte Verbesserung der Güte der radiologischen R0/R1-Diagnose bei Verwendung eines Versatzes in den Randabstandskriterien von ca. 2 mm sollte in zukünftigen Studien mit Blick auf die mögliche praxisnahe klinische Relevanz näher untersucht werden.

Obwohl die DBT robuster gegenüber unterschiedlichen Readern ist, kann auf Basis der Ergebnisse dieser Arbeit und unter Einbeziehung der weiteren vorliegenden Forschungsergebnisse nicht allgemeingültig geschlossen werden, dass durch die DBT eine Verbesserung im intraoperativen Einsatz gegenüber der PR zu erzielen ist.

Da die DBT jedoch eine höhere Inter-Reader-Reliabilität aufweist und Herdbefundgrößen gegenüber der PR zuverlässiger ermittelt werden können, legt dies

eine mögliche Verbesserung beim präoperativen Einsatz nahe. Die Übertragbarkeit der Ergebnisse bei der Diagnose von Präparaten auf die Mammakarzinomdiagnostik sowie die klinische Relevanz der erzielbaren Verbesserung sollten daher Gegenstand weiterführender Studien sein.

Im Rahmen der Diskussion in dieser Arbeit wurde die korrekte Identifikation des Ortes des kleinsten Randabstandes als einer der möglicherweise größten Einflussfaktoren auf die Güte der Vorhersage des histologischen Ergebnisses identifiziert. Im Gebiet des präoperativen Einsatzes geben erste Studien zu KI-Methoden für einzelne radiologische Aufgaben die Indikation, dass diese gegenüber herkömmlichen computergestützten Algorithmen und teilweise auch gegenüber Vergleichsgruppen aus Radiologinnen bessere Ergebnisse erzielen können (siehe z.B. [101, 102]). Auch wenn dies noch durch weitere Studien nachgewiesen werden muss und KI-Methoden nicht uneingeschränkt als universelle Lösung betrachtet werden dürfen [103, 104], könnten sie in dem vorliegenden spezifischen Anwendungsfall dennoch dazu beitragen, die intraoperative Wertigkeit der radiologischen Methoden und vor allem die Voraussage des histologischen Ergebnisses zu verbessern. Sei es durch Verbesserung der Bildqualität, eine umfangreichere und automatisierte Randabstandsbestimmung oder bei der Bestimmung einer „digitalen Zweitmeinung“ für die R0/R1-Diagnose. Unterstützt von der weiteren Entwicklung im Bereich der Rechenleistung sowie Datenspeicherung und -verarbeitung stellt dies ein vielversprechendes Forschungsfeld für die Zukunft dar, um das Patientinnenwohl weiter zu steigern.

6 Literatur

1. Perry N., Broeders M., de Wolf C. et al. (Hrsg): *European guidelines for quality assurance in breast cancer screening and diagnosis*, 4. Aufl. Off. for Official Publ. of the Europ. Communities, Luxembourg, 2006
2. Tiling R., Avril N., Lissner J., Hahn K. (Hrsg): *Mammakarzinom. Nuklearmedizinische und radiologische Diagnostik*. Springer, Berlin, 1998
3. Erdmann F., Spix C., Katalinic A. et al.: *Krebs in Deutschland für 2017/2018*. Robert Koch-Institut, 2021
4. Horn J., Åsvold B. O., Opdahl S., Tretli S., Vatten L. J.: *Reproductive factors and the risk of breast cancer in old age: a Norwegian cohort study*. *Breast Cancer Res Treat* 139 (Nr. 1): S. 237–243, 2013. doi:10.1007/s10549-013-2531-0
5. Böcker W., Denk H., Heitz P. U., Höfler G., Kreipe H. H., Moch H. (Hrsg): *Pathologie*, 5. Aufl. Elsevier Urban & Fischer, München, 2012
6. Hauner D., Janni W., Rack B., Hauner H.: *The effect of overweight and nutrition on prognosis in breast cancer*. *Dtsch Arztebl Int* 108 (Nr. 47): S. 795–801, 2011. doi:10.3238/arztebl.2011.0795
7. Dall G., Risbridger G., Britt K.: *Mammary stem cells and parity-induced breast cancer protection- new insights*. *J Steroid Biochem Mol Biol* 170: S. 54–60, 2017. doi:10.1016/j.jsbmb.2016.02.018
8. Leitlinienprogramm Onkologie (Deutsche Krebsgesellschaft, Deutsche Krebshilfe, AWMF): *S3-Leitlinie Früherkennung, Diagnose, Therapie und Nachsorge des Mammakarzinoms, Version 4.2, AWMF Registernummer: 032-045OL*. <http://www.leitlinienprogramm-onkologie.de/leitlinien/mammakarzinom/>. Zugegriffen: 01. Mai 2021
9. Robert Koch-Institut und die Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V.: *Krebs in Deutschland 2015/2016*, Berlin, 2019
10. Barnes B., Kraywinkel K., Nowossadeck E. et al.: *Bericht zum Krebsgeschehen in Deutschland 2016*. Zentrum für Krebsregisterdaten im Robert Koch-Institut, 2016
11. Petru E., Fink D., Köchli O. R., Loibl S. (Hrsg): *Praxisbuch Gynäkologische Onkologie*, 5. Aufl. Springer, Berlin, Heidelberg, 2019

12. Senkus E., Kyriakides S., Ohno S. et al.: *Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up*. Ann Oncol 26 Suppl 5: v8-30, 2015. doi:10.1093/annonc/mdv298
13. Inwald E. C., Ortmann O.: *Nachsorge nach gynäkologischen Malignomen und Mammakarzinom*. Gynäkologe 54 (Nr. 4): S. 281–290, 2021. doi:10.1007/s00129-021-04767-8
14. Haberstroh M., Heitz F., Barinoff J. et al.: *Der Einfluss des lokoregionären Rezidivs (LR) auf das Gesamtüberleben beim Mammakarzinom*. Senologie - Zeitschrift für Mammadiagnostik und -therapie 6 (Nr. 02), 2009. doi:10.1055/s-0029-1224966
15. Dijkshoorn A. B. C., van Stralen H. E., Sloots M., Schagen S. B., Visser-Meily J. M. A., Schepers V. P. M.: *Prevalence of cognitive impairment and change in patients with breast cancer: A systematic review of longitudinal studies*. Psychooncology 30 (Nr. 5): S. 635–648, 2021. doi:10.1002/pon.5623
16. Okamura M., Yamawaki S., Akechi T., Taniguchi K., Uchitomi Y.: *Psychiatric disorders following first breast cancer recurrence: prevalence, associated factors and relationship to quality of life*. Jpn J Clin Oncol 35 (Nr. 6): S. 302–309, 2005. doi:10.1093/jjco/hyi097
17. Okamura H., Yamamoto N., Watanabe T. et al.: *Patients' understanding of their own disease and survival potential in patients with metastatic breast cancer*. Breast Cancer Res Treat 61 (Nr. 2): S. 145–150, 2000. doi:10.1023/A:1006483214678
18. Lokuhetty D., White V. A., Watanabe R., Cree I. A.: *Breast tumours*, 5. Aufl. World Health Organization Classification of tumours. International Agency for Research on Cancer, Lyon, 2019
19. Käab-Sanyal V.: *Jahresbericht Evaluation 2017 Deutsches Mammographie-Screening Programm*, Berlin, 2019
20. Mannu G. S., Wang Z., Broggio J. et al.: *Invasive breast cancer and breast cancer mortality after ductal carcinoma in situ in women attending for breast screening in England, 1988-2014: population based observational cohort study*. BMJ 369: m1570, 2020. doi:10.1136/bmj.m1570

21. Yen T. W. F., Hunt K. K., Ross M. I. et al.: *Predictors of invasive breast cancer in patients with an initial diagnosis of ductal carcinoma in situ: a guide to selective use of sentinel lymph node biopsy in management of ductal carcinoma in situ*. J Am Coll Surg 200 (Nr. 4): S. 516–526, 2005. doi:10.1016/j.jamcollsurg.2004.11.012.
22. Epstein M. S., Lagios M. D., Silverstein M. J.: *Ductal Carcinoma in Situ of the Breast* The Breast, 5. Aufl. Elsevier, S. 562–575, 2018
23. Brierley J. D.: *TNM Classification of Malignant Tumours*, 8. Aufl. John Wiley & Sons Incorporated, Chicester, 2017
24. Fansa H., Heitmann C.: *Brustchirurgie*. Springer, Berlin, Heidelberg, 2018
25. Barkhausen J., Rody A., Schäfer F. K. W.: *Digitale Tomosynthese der Brust*. Thieme, Stuttgart, 2015
26. Reiser M., Kuhn F.-P., Debus J.: *Radiologie*. Thieme, Stuttgart, 2017
27. *ACR BI-RADS®-Atlas der Mammadiagnostik*. Springer, Berlin, Heidelberg, 2016
28. Fischer U. (Hrsg): *ACR BI-RADS*. Illustrierte Anleitung zur einheitlichen Befunderstellung von Mammographie, Mammasonographie, MR-Mammographie, 2. Aufl. Thieme, Stuttgart, 2006
29. Fisher R. A.: *Statistical methods for research workers*, 12. Aufl. Biological monographs and manuals. Oliver and Boyd, Edinburgh u.a., 1954
30. Heywang-Köbrunner S. H., Schreer I.: *Bildgebende Mammadiagnostik*. Thieme, Stuttgart, 2015
31. Fischer U., Baum F.: *Mamma*. Pareto-Reihe Radiologie. Thieme, Stuttgart, 2007
32. Fischer U., Baum F.: *Diagnostik und Therapie des Mammakarzinoms*. Thieme, Stuttgart, 2014
33. Yuan W.-H., Hsu H.-C., Chen Y.-Y., Wu C.-H.: *Supplemental breast cancer-screening ultrasonography in women with dense breasts: a systematic review and meta-analysis*. Br J Cancer 123 (Nr. 4): S. 673–688, 2020. doi:10.1038/s41416-020-0928-1
34. Wanders J. O. P., Holland K., Veldhuis W. B. et al.: *Volumetric breast density affects performance of digital screening mammography*. Breast

- Cancer Res Treat 162 (Nr. 1): S. 95–103, 2017. doi:10.1007/s10549-016-4090-7
35. Preibsch H., Siegmann-Luz K. C.: *Digitale Tomosynthese der Mamma*. Radiologe 55 (Nr. 1): 59-67; quiz 68-70, 2015. doi:10.1007/s00117-014-2753-0
36. Bani M. R., Lux M. P., Heusinger K. et al.: *Factors correlating with reexcision after breast-conserving therapy*. Eur J Surg Oncol 35 (Nr. 1): S. 32–37, 2009. doi:10.1016/j.ejso.2008.04.008
37. Andersson I., Ikeda D. M., Zackrisson S. et al.: *Breast tomosynthesis and digital mammography: a comparison of breast cancer visibility and BI-RADS classification in a population of cancers with subtle mammographic findings*. Eur Radiol 18 (Nr. 12): S. 2817–2825, 2008. doi:10.1007/s00330-008-1076-9
38. Kontos D., Bakic P. R., Carton A.-K., Troxel A. B., Conant E. F., Maidment A. D. A.: *Parenchymal texture analysis in digital breast tomosynthesis for breast cancer risk estimation: a preliminary study*. Acad Radiol 16 (Nr. 3): S. 283–298, 2009. doi:10.1016/j.acra.2008.08.014
39. Ciatto S., Houssami N., Bernardi D. et al.: *Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study*. The Lancet Oncology 14 (Nr. 7): S. 583–589, 2013. doi:10.1016/S1470-2045(13)70134-7
40. Friedewald S. M., Rafferty E. A., Rose S. L. et al.: *Breast cancer screening using tomosynthesis in combination with digital mammography*. JAMA 311 (Nr. 24): S. 2499–2507, 2014. doi:10.1001/jama.2014.6095
41. Scherz N.: *Tomosynthese kombiniert mit Mammographie entdeckt mehr Mammakarzinome als Mammographie allein*. Praxis 103 (Nr. 20): S. 1213–1214, 2014. doi:10.1024/1661-8157/a001797
42. Fontaine M., Tourasse C., Pages E. et al.: *Local Tumor Staging of Breast Cancer: Digital Mammography versus Digital Mammography Plus Tomosynthesis*. Radiology 291 (Nr. 3): S. 594–603, 2019. doi:10.1148/radiol.2019182457

43. Wallis M. G., Moa E., Zanca F., Leifland K., Danielsson M.: *Two-view and single-view tomosynthesis versus full-field digital mammography: high-resolution X-ray imaging observer study*. *Radiology* 262 (Nr. 3): S. 788–796, 2012. doi:10.1148/radiol.11103514
44. Teertstra H. J., Loo C. E., van den Bosch M. A. A. J. et al.: *Breast tomosynthesis in clinical practice: initial results*. *Eur Radiol* 20 (Nr. 1): S. 16–24, 2010. doi:10.1007/s00330-009-1523-2
45. Michell M. J., Wasan R. K., Whelehan P. et al.: *Digital breast tomosynthesis: a comparison of the accuracy of digital breast tomosynthesis, two-dimensional digital mammography and two-dimensional screening mammography (film-screen)*. *Breast Cancer Res* 11 Suppl 2: O1-6, P1-33, 2009. doi:10.1186/bcr2365
46. Gur D., Abrams G. S., Chough D. M. et al.: *Digital breast tomosynthesis: observer performance study*. *AJR Am J Roentgenol* 193 (Nr. 2): S. 586–591, 2009. doi:10.2214/AJR.08.2031
47. Ott V. E. M.: *Diagnostische Wertigkeit der Tomosynthese (3D-Mammographie) als Ergänzung zur digitalen 2D-Mammographie in der Abklärung von Befunden der Kategorie BI-RADS 3, 4 und 5*. Inaugural-Dissertation, Eberhard Karls Universität zu Tübingen, 2015
48. Zuley M. L., Guo B., Catullo V. J. et al.: *Comparison of two-dimensional synthesized mammograms versus original digital mammograms alone and in combination with tomosynthesis images*. *Radiology* 271 (Nr. 3): S. 664–671, 2014. doi:10.1148/radiol.13131530
49. Heywang-Köbrunner S. H., Jänsch A., Hacker A., Weinand S., Vogelmann T.: *Digital breast tomosynthesis (DBT) plus synthesised two-dimensional mammography (s2D) in breast cancer screening is associated with higher cancer detection and lower recalls compared to digital mammography (DM) alone: results of a systematic review and meta-analysis*. *Eur Radiol* 32 (Nr. 4): S. 2301–2312, 2022. doi:10.1007/s00330-021-08308-8
50. Tucker A. W., Calliste J., Gidcumb E. M. et al.: *Comparison of a stationary digital breast tomosynthesis system to magnified 2D mammography using*

- breast tissue specimens*. Acad Radiol 21 (Nr. 12): S. 1547–1552, 2014. doi:10.1016/j.acra.2014.07.009
51. Förnvik D., Zackrisson S., Ljungberg O. et al.: *Breast tomosynthesis: Accuracy of tumor measurement compared with digital mammography and ultrasonography*. Acta Radiol 51 (Nr. 3): S. 240–247, 2010. doi:10.3109/02841850903524447
52. Urano M., Shiraki N., Kawai T. et al.: *Digital mammography versus digital breast tomosynthesis for detection of breast cancer in the intraoperative specimen during breast-conserving surgery*. Breast Cancer 23 (Nr. 5): S. 706–711, 2016. doi:10.1007/s12282-015-0628-5
53. Schulz-Wendtland R., Dilbat G., Bani M. R. et al.: *Use of Tomosynthesis in Intraoperative Digital Specimen Radiography - Is a Reduction of Breast Re-excision Rates Possible?* Geburtshilfe Frauenheilkd 71 (Nr. 12): S. 1080–1084, 2011. doi:10.1055/s-0031-1280427
54. Schulz-Wendtland R., Dilbat G., Bani M. et al.: *Full Field Digital Mammography (FFDM) versus CMOS Technology versus Tomosynthesis (DBT) - Which System Increases the Quality of Intraoperative Imaging?* Geburtshilfe Frauenheilkd 72 (Nr. 6): S. 532–538, 2012. doi:10.1055/s-0032-1314942
55. Schulz-Wendtland R., Dilbat G., Bani M. et al.: *Full Field Digital Mammography (FFDM) versus CMOS Technology, Specimen Radiography System (SRS) and Tomosynthesis (DBT) - Which System Can Optimise Surgical Therapy?* Geburtshilfe Frauenheilkd 73 (Nr. 5): S. 422–427, 2013. doi:10.1055/s-0032-1328600
56. Tingberg A., Förnvik D., Mattsson S., Svahn T., Timberg P., Zackrisson S.: *Breast cancer screening with tomosynthesis--initial experiences*. Radiat Prot Dosimetry 147 (1-2): S. 180–183, 2011. doi:10.1093/rpd/ncr296
57. Delorme S., Debus J., Jenderka K.-V.: *Sonografie*. Thieme, Stuttgart, 2012
58. Fischer U.: *Atlas der MR-Mammografie*. Hochauflösende Mamma-MRT ; 25 Tabellen, 2. Aufl. Thieme, Stuttgart, 2010

59. Deutsche Krebsgesellschaft e.V., Deutsche Gesellschaft für Senologie e.V., Zertifizierungskommission Brustkrebszentren: *Jahresbericht der zertifizierten Brustkrebszentren. Auditjahr 2019/ Kennzahlenjahr 2018, 2020*
60. European comission: *Surgical guidelines for the management of breast cancer*. Eur J Surg Oncol 35 Suppl 1: S. 1–22, 2009.
doi:10.1016/j.ejso.2009.01.008
61. Fisher B., Anderson S., Bryant J. et al.: *Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer*. N Engl J Med 347 (Nr. 16): S. 1233–1241, 2002. doi:10.1056/NEJMoa022152
62. Zehra S., Doyle F., Barry M., Walsh S., Kell M. R.: *Health-related quality of life following breast reconstruction compared to total mastectomy and breast-conserving surgery among breast cancer survivors: a systematic review and meta-analysis*. Breast Cancer 27 (Nr. 4): S. 534–566, 2020.
doi:10.1007/s12282-020-01076-1
63. Moran M. S., Schnitt S. J., Giuliano A. E. et al.: *Society of Surgical Oncology-American Society for Radiation Oncology consensus guideline on margins for breast-conserving surgery with whole-breast irradiation in stages I and II invasive breast cancer*. J Clin Oncol 32 (Nr. 14): S. 1507–1515, 2014. doi:10.1200/JCO.2013.53.3935
64. Dillon M. F., Hill A. D. K., Quinn C. M., McDermott E. W., O'Higgins N.: *A pathologic assessment of adequate margin status in breast-conserving therapy*. Ann Surg Oncol 13 (Nr. 3): S. 333–339, 2006.
doi:10.1245/ASO.2006.03.098
65. Morrow M., van Zee K. J., Solin L. J. et al.: *Society of Surgical Oncology-American Society for Radiation Oncology-American Society of Clinical Oncology Consensus Guideline on Margins for Breast-Conserving Surgery With Whole-Breast Irradiation in Ductal Carcinoma in Situ*. Pract Radiat Oncol 6 (Nr. 5): S. 287–295, 2016. doi:10.1016/j.prro.2016.06.011
66. Morrow M., Jagsi R., Alderman A. K. et al.: *Surgeon recommendations and receipt of mastectomy for treatment of breast cancer*. JAMA 302 (Nr. 14): S. 1551–1556, 2009. doi:10.1001/jama.2009.1450

67. Havel L., Naik H., Ramirez L., Morrow M., Landercasper J.: *Impact of the SSO-ASTRO Margin Guideline on Rates of Re-excision After Lumpectomy for Breast Cancer: A Meta-analysis*. Ann Surg Oncol 26 (Nr. 5): S. 1238–1244, 2019. doi:10.1245/s10434-019-07247-5
68. McCahill L. E., Single R. M., Aiello Bowles E. J. et al.: *Variability in reexcision following breast conservation surgery*. JAMA 307 (Nr. 5): S. 467–475, 2012. doi:10.1001/jama.2012.43
69. Mamtani A., Romanoff A., Baser R., Vincent A., Morrow M., Gemignani M. L.: *Adoption of SSO-ASTRO Margin Guidelines for Ductal Carcinoma in Situ: What Is the Impact on Use of Additional Surgery?* Ann Surg Oncol 28 (Nr. 1): S. 295–302, 2021. doi:10.1245/s10434-020-08610-7
70. Britton P. D., Sonoda L. I., Yamamoto A. K., Koo B., Soh E., Goud A.: *Breast surgical specimen radiographs: how reliable are they?* Eur J Radiol 79 (Nr. 2): S. 245–249, 2011. doi:10.1016/j.ejrad.2010.02.012
71. Angarita F. A., Nadler A., Zerhouni S., Escallon J.: *Perioperative measures to optimize margin clearance in breast conserving surgery*. Surg Oncol 23 (Nr. 2): S. 81–91, 2014. doi:10.1016/j.suronc.2014.03.002
72. Thill M., Barinoff J., Hoellen F., Faridi A.: *Der Resektionsrand und seine Beurteilung*. Gynäkologe 49 (Nr. 3): S. 152–158, 2016. doi:10.1007/s00129-015-3798-5
73. Gray R. J., Pockaj B. A., Garvey E., Blair S.: *Intraoperative Margin Management in Breast-Conserving Surgery: A Systematic Review of the Literature*. Ann Surg Oncol 25 (Nr. 1): S. 18–27, 2018. doi:10.1245/s10434-016-5756-4
74. St John E. R., Al-Khudairi R., Ashrafian H. et al.: *Diagnostic Accuracy of Intraoperative Techniques for Margin Assessment in Breast Cancer Surgery: A Meta-analysis*. Ann Surg 265 (Nr. 2): S. 300–310, 2017. doi:10.1097/SLA.0000000000001897
75. Schmachtenberg C., Engelken F., Fischer T., Bick U., Poellinger A., Fallenberg E. M.: *Intraoperative specimen radiography in patients with nonpalpable malignant breast lesions*. Rofo 184 (Nr. 7): S. 635–642, 2012. doi:10.1055/s-0032-1312730

76. Chagpar A. B., Butler M., Killelea B. K., Horowitz N. R., Stavris K., Lannin D. R.: *Does three-dimensional intraoperative specimen imaging reduce the need for re-excision in breast cancer patients? A prospective cohort study.* Am J Surg 210 (Nr. 5): S. 886–890, 2015. doi:10.1016/j.amjsurg.2015.05.018
77. Holland R., Schuurmans Stekhoven J. H., Hendriks J., Verbeek A., Mravunac M.: *Extent, distribution, and mammographic/histological correlations of breast ductal carcinoma in situ.* The Lancet 335 (Nr. 8688): S. 519–522, 1990. doi:10.1016/0140-6736(90)90747-S
78. Poellinger A., Diekmann S., Dietz E., Bick U., Diekmann F.: *In patients with DCIS: is it sufficient to histologically examine only those tissue specimens that contain microcalcifications?* Eur Radiol 18 (Nr. 5): S. 925–930, 2008. doi:10.1007/s00330-007-0846-0
79. Graham R. A., Homer M. J., Katz J., Rothschild J., Safaii H., Supran S.: *The pancake phenomenon contributes to the inaccuracy of margin assessment in patients with breast cancer.* The American Journal of Surgery 184 (Nr. 2): S. 89–93, 2002. doi:10.1016/S0002-9610(02)00902-9
80. Grunert J.-H., Wiechmann U., Gmelin E.: *Kompressionsfreie Mammapräparateradiographie in zwei Ebenen mit Hilfe eines Radiographiecontainers.* Rofo 175 (Nr. 3): S. 413–417, 2003. doi:10.1055/s-2003-37821
81. Rua C., Lebas P., Michenet P., Ouldamer L.: *Evaluation of lumpectomy surgical specimen radiographs in subclinical, in situ and invasive breast cancer, and factors predicting positive margins.* Diagn Interv Imaging 93 (Nr. 11): S. 871–877, 2012. doi:10.1016/j.diii.2012.07.010
82. Mario J., Venkataraman S., Fein-Zachary V., Knox M., Brook A., Slanetz P.: *Lumpectomy Specimen Radiography: Does Orientation or 3-Dimensional Tomosynthesis Improve Margin Assessment?* Can Assoc Radiol J 70 (Nr. 3): S. 282–291, 2019. doi:10.1016/j.carj.2019.03.005
83. Amer H.: *Specimen radiography: digital breast tomosynthesis versus full field digital mammography : which modality provides more accurate prediction of margin status?* Dissertation, Charité - Universitätsmedizin Berlin

84. Passing H., Bablok: *A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in clinical chemistry, Part I.* J Clin Chem Clin Biochem 21 (Nr. 11): S. 709–720, 1983.
doi:10.1515/cclm.1983.21.11.709
85. Bland J. M., Altman D. G.: *Statistical methods for assessing agreement between two methods of clinical measurement.* The Lancet 1 (Nr. 8476): S. 307–310, 1986
86. Altman D. G.: *Practical Statistics for Medical Research.* Chapman and Hall/CRC Texts in Statistical Science. CRC Press LLC, Boca Raton, 1990
87. Cohen J.: *A Coefficient of Agreement for Nominal Scales.* Educational and Psychological Measurement 20 (Nr. 1): S. 37–46, 1960.
doi:10.1177/001316446002000104
88. Krebsregister in Baden-Württemberg: *Krebs in Baden-Württemberg.* Kurzbericht für die Jahre 2014/2015. Korrigierte Version 15.06.2020.
Datenstand: 15.10.2018. https://www.krebsregister-bw.de/fileadmin/user_upload/filemount/allgemein/Publikationen/EKR-Kurzbericht_2014-2015_idF_150620.pdf. Zugegriffen: 02. Mai 2021
89. Weigel S., Heindel W., Dietz C., Meyer-Johann U., Graewingholt A., Hense H. W.: *Stratifizierung des Brustkrebsrisikos hinsichtlich der Einflüsse von Alter und mammografischer Dichte.* Rofo 192 (Nr. 7): S. 678–685, 2020. doi:10.1055/a-1100-0016
90. Galati F., Marzocca F., Bassetti E. et al.: *Added Value of Digital Breast Tomosynthesis Combined with Digital Mammography According to Reader Agreement: Changes in BI-RADS Rate and Follow-Up Management.* Breast Care (Basel) 12 (Nr. 4): S. 218–222, 2017. doi:10.1159/000477537
91. Basha M. A. A., Safwat H. K., Alaa Eldin A. M., Dawoud H. A., Hassanin A. M.: *The added value of digital breast tomosynthesis in improving diagnostic performance of BI-RADS categorization of mammographically indeterminate breast lesions.* Insights Imaging 11 (Nr. 1): S. 26, 2020.
doi:10.1186/s13244-020-0835-2

92. Ross J. S., Delamarter R., Hueftle M. G. et al.: *Gadolinium-DTPA-enhanced MR imaging of the postoperative lumbar spine: time course and mechanism of enhancement*. AJR Am J Roentgenol 152 (Nr. 4): S. 825–834, 1989. doi:10.2214/ajr.152.4.825
93. Romanucci G., Mercogliano S., Carucci E. et al.: *Diagnostic accuracy of resection margin in specimen radiography: digital breast tomosynthesis versus full-field digital mammography*. Radiol Med, 2021. doi:10.1007/s11547-021-01337-9
94. Seo N., Kim H. H., Shin H. J. et al.: *Digital breast tomosynthesis versus full-field digital mammography: comparison of the accuracy of lesion measurement and characterization using specimens*. Acta Radiol 55 (Nr. 6): S. 661–667, 2014. doi:10.1177/0284185113503636
95. Amer H. A., Schmitzberger F., Ingold-Heppner B. et al.: *Digital breast tomosynthesis versus full-field digital mammography-Which modality provides more accurate prediction of margin status in specimen radiography?* Eur J Radiol 93: S. 258–264, 2017. doi:10.1016/j.ejrad.2017.05.041
96. Mazouni C., Rouzier R., Balleyguier C. et al.: *Specimen radiography as predictor of resection margin status in non-palpable breast lesions*. Clin Radiol 61 (Nr. 9): S. 789–796, 2006. doi:10.1016/j.crad.2006.04.017
97. Coombs N. J., Vassallo P. P., Parker A. J., Yiangou C.: *Radiological review of specimen radiographs after breast localisation biopsy is not always necessary*. Eur J Surg Oncol 32 (Nr. 5): S. 516–519, 2006. doi:10.1016/j.ejso.2006.02.019
98. McCormick J. T., Keleher A. J., Tikhomirov V. B., Budway R. J., Caushaj P. F.: *Analysis of the use of specimen mammography in breast conservation therapy*. The American Journal of Surgery 188 (Nr. 4): S. 433–436, 2004. doi:10.1016/j.amjsurg.2004.06.030
99. Bimston D. N., Bebb G. G., Wagman L. D.: *Is specimen mammography beneficial?* Arch Surg 135 (Nr. 9): 1083-6; discussion 1086-9, 2000. doi:10.1001/archsurg.135.9.1083
100. Partain N., Calvo C., Mokdad A. et al.: *Differences in Re-excision Rates for Breast-Conserving Surgery Using Intraoperative 2D Versus 3D*

- Tomosynthesis Specimen Radiograph*. Ann Surg Oncol 27 (Nr. 12): S. 4767–4776, 2020. doi:10.1245/s10434-020-08877-w
101. Wu N., Phang J., Park J. et al.: *Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening*. IEEE Transactions on Medical Imaging 39 (Nr. 4): S. 1184–1194, 2020. doi:10.1109/TMI.2019.2945514
102. Conant E. F., Toledano A. Y., Periaswamy S. et al.: *Improving Accuracy and Efficiency with Concurrent Use of Artificial Intelligence for Digital Breast Tomosynthesis*. Radiol Artif Intell 1 (Nr. 4): e180096, 2019. doi:10.1148/ryai.2019180096
103. Geras K. J., Mann R. M., Moy L.: *Artificial Intelligence for Mammography and Digital Breast Tomosynthesis: Current Concepts and Future Perspectives*. Radiology 293 (Nr. 2): S. 246–259, 2019. doi:10.1148/radiol.2019182627
104. Baltzer P. A. T.: *Künstliche Intelligenz in der Mammadiagnostik : Anwendungsgebiete aus klinischer Perspektive*. Radiologe 61 (Nr. 2): S. 192–198, 2021. doi:10.1007/s00117-020-00802-2

7 Erklärung zum Eigenanteil

Die Arbeit wurde in der Klinik für Interventionelle und Diagnostische Radiologie der Uniklinik Tübingen unter der Betreuung von Frau PD Dr. med. Heike Preibsch realisiert.

Die Konzeption der Studie erfolgte in Zusammenarbeit mit Frau PD Dr. med. Heike Preibsch und Frau Dr. med. Sabrina Fleischer.

Die Ausarbeitung und Erstellung der Bewertungsbögen sowie die Dokumentation der Daten erfolgte durch mich, Ramona Walder.

Die Auswertung der präoperativen Mammographien sowie der intraoperativen Tomosynthese und Präparateradiographien anhand der Bewertungsbögen, wurde von Frau PD Dr. med. Heike Preibsch und Frau Dr.med. Sabrina Fleischer umgesetzt.

Die statistische Auswertung und Auswahl der zu verwendenden statistischen Methoden wurde von mir, Ramona Walder, mit Unterstützung von Dr.-Ing. Thorben Walder durchgeführt. Eine statistische Beratung erfolgte nach Ausarbeitung aller Ergebnisse zur Rückversicherung bezüglich der korrekten Anwendung der gewählten Methoden durch Dr. Johann Jacoby vom Institut für Biometrie.

Ich versichere, das Manuskript selbständig verfasst zu haben und keine weiteren als die von mir, Ramona Walder, angegebenen Quellen verwendet zu haben.

Tübingen, den 26.06.2023

8 Anhang

8.1 Bewertungsbogen Mammographie

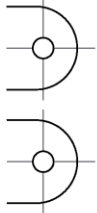
Bewertungsbogen präoperative Mammographie	Bewertungsbogen präoperative Mammographie
<p>Kopfdaten RIS-Nummer: _____ bei _____ Uhr _____ cm MA Gesichertes Mammakarzinom _____</p> <p>1. Einteilung Entität <input type="checkbox"/> Duktales Karzinom <input type="checkbox"/> DCIS <input type="checkbox"/> Lobuläres Karzinom <input type="checkbox"/> Nicht definierbar</p> <p>2. Herdbefund</p> <p>2.1 Lokalisation Bitte Lage ankreuzen:  rechte Brust linke Brust</p> <p>2.2 Größe lateromedial craniokaudal CCx _____ [mm] MLOy _____ [mm] ventrodorsal CCz _____ [mm] MLOz _____ [mm]</p> <p>2.3 Form <input type="checkbox"/> oval <input type="checkbox"/> miklobulliert <input type="checkbox"/> umschrieben <input type="checkbox"/> rund <input type="checkbox"/> unscharf <input type="checkbox"/> überlagert <input type="checkbox"/> irregulär <input type="checkbox"/> spikuliert <input type="checkbox"/> Nicht abgrenzbar</p> <p>3. suspekt Mikrokalzifikationen vorhanden: <input type="checkbox"/> ja <input type="checkbox"/> nein</p> <p>3.1 Lokalisation <input type="checkbox"/> Verkalkungen in Herdläsion <input type="checkbox"/> Verkalkungen außerhalb Herdläsion</p> <p>3.2 Größe lateromedial craniokaudal CCx _____ [mm] MLOy _____ [mm] ventrodorsal CCz _____ [mm] MLOz _____ [mm]</p> <p>4. Brustdicke nach ACR <input type="checkbox"/> ACR A <input type="checkbox"/> ACR B <input type="checkbox"/> ACR C <input type="checkbox"/> ACR D</p> <p>5. Technische Daten Brustdicke nach Kompression _____ [mm] Kraft auf Brust _____ [N]</p> <p style="text-align: right;">Seite 1</p>	<p>6. Subjektive Einschätzung</p> <p>6.1 Wie gut konnten Sie den Herdbefund erkennen und messen? <input type="checkbox"/> Sehr schlecht <input type="checkbox"/> schlecht <input type="checkbox"/> mittel <input type="checkbox"/> gut <input type="checkbox"/> Sehr gut</p> <p>6.2 Wie gut konnten Sie den Mikrokalz erkennen und messen? <input type="checkbox"/> Sehr schlecht <input type="checkbox"/> schlecht <input type="checkbox"/> mittel <input type="checkbox"/> gut <input type="checkbox"/> Sehr gut</p> <p>6.3 Wie empfinden Sie den zeitlichen Aufwand der Messerhebungen? <input type="checkbox"/> Sehr hoch <input type="checkbox"/> hoch <input type="checkbox"/> mittel <input type="checkbox"/> gering <input type="checkbox"/> Sehr gering</p> <p>7. Bearbeitungsdauer Zeit gestoppt: _____ [min]</p> <p style="text-align: right;">Seite 2</p>

Abbildung 8.1: Bewertungsbogen Mammographie

8.2 Bewertungsbogen intraoperative PR

Bewertungsbogen postoperative Präparat-Radiographie

Kopfdaten
 RIS-Nummer: _____ bei _____ Uhr _____ cm MA
 Gesichertes Mammarkarzinom

1. Präparat Größe
 APx _____ [mm] Lly _____ [mm]
 APz _____ [mm] LLz _____ [mm]

2. Herdbefund

2.1 Größe
 APx _____ [mm] Lly _____ [mm]
 APz _____ [mm] LLz _____ [mm]

2.2 Form
 oval
 rund
 irregulär

2.3 Begrenzung
 mikrobulliert
 unregelmäßig
 spikuliert

umschrieben
 überlagert

Bewertungsbogen postoperative Präparat-Radiographie

R0* R1

* R0 \approx 10mm Abstand invasiv
 R0 \approx 2mm Abstand Mikrokalz

5. Technische Daten
 Dicke Präparat _____ [mm] Kraft auf Präparat _____ [N]

6. Subjektive Einschätzung

6.1 Wie gut konnten Sie den Herdbefund erkennen und messen?
 Sehr schlecht schlecht mittel gut Sehr gut

6.2 Wie gut konnten Sie den Mikrokalz erkennen und messen?
 Sehr schlecht schlecht mittel gut Sehr gut

6.3 Wie empfinden Sie den zeitlichen Aufwand der Messerhebungen?
 Sehr hoch hoch mittel gering Sehr gering

7. Bearbeitungsdauer
 Zeit gestoppt: _____ [min]

Bewertungsbogen postoperative Präparat-Radiographie

3.1 Lokalisation
 Verkalkungen in Herdläsion
 Verkalkungen außerhalb Herdläsion

Nachresektion empfohlen: ja nein

3.2 Größe
 APx _____ [mm] Lly _____ [mm]
 APz _____ [mm] LLz _____ [mm]

4. Randabstand

mammilär	_____ TU/MK [mm]	LL	mammilär	_____ TU/MK [mm]
peripher	_____ TU/MK [mm]		peripher	_____ TU/MK [mm]
kranial	_____ TU/MK [mm]		kutan	_____ TU/MK [mm]
kaudal	_____ TU/MK [mm]		dorsal	_____ TU/MK [mm]
medial	_____ TU/MK [mm]			
lateral	_____ TU/MK [mm]			

Bewertungsbogen postoperative Präparat-Radiographie

Seite 2

Abbildung 8.2: Bewertungsbogen intraoperative PR

8.3 Bewertungsbogen intraoperative DBT

Bewertungsbogen postoperative Präparat-Tomosynthese

Kopfdaten
RIS-Nummer: _____ bei _____ Uhr _____ cm MA
Gesichertes Mammarkarzinom _____

1. Präparat Größe
APx _____ [mm] Lly _____ [mm]
APz _____ [mm] LLz _____ [mm]

2. Herdbefund

2.1 Größe
APx _____ [mm] Lly _____ [mm]
APz _____ [mm] LLz _____ [mm]

2.2 Form
 oval miklobulliert ungeschrieben
 rund unscharf überlagert
 irregulär spikuliert

3. suspekte Mikrokalzifikationen vorhanden: ja nein

3.1 Lokalisation
 Verkalkungen in Herdläsion APx _____ [mm] Lly _____ [mm]
 Verkalkungen außerhalb Herdläsion APz _____ [mm] LLz _____ [mm]

Nachsrektion empfohlen: ja nein

4. Randabstand
AP mammilär _____ TUMK[mm] LL mammilär _____ TUMK[mm]
peripher _____ TUMK[mm] peripher _____ TUMK[mm]
kranial _____ TUMK[mm] kranial _____ TUMK[mm]
kaudal _____ TUMK[mm] kaudal _____ TUMK[mm]
medial _____ TUMK[mm] medial _____ TUMK[mm]
lateral _____ TUMK[mm] lateral _____ TUMK[mm]

Bewertungsbogen postoperative Präparat-Tomosynthese

R0* R1

R0 10mm Abstand invasiv
R0 2mm Abstand Mikrokalz

5. Technische Daten
Dicke Präparat _____ [mm] Kraft auf Präparat _____ [N]

6. Subjektive Einschätzung

6.1 Wie gut konnten Sie den Herdbefund erkennen und messen?
 Sehr schlecht schlecht mittel gut Sehr gut

6.2 Wie gut konnten Sie den Mikrokalz erkennen und messen?
 Sehr schlecht schlecht mittel gut Sehr gut

6.3 Wie empfinden Sie den zeitlichen Aufwand der Messerhebungen?
 Sehr hoch hoch mittel gering Sehr gering

7. Bearbeitungsdauer
Zeit gestoppt: _____ [min]

Seite 1

Seite 2

Abbildung 8.3: Bewertungsbogen intraoperative DBT

8.4 Weitere Diagramme Inter-Reader-Vergleich Mammographie

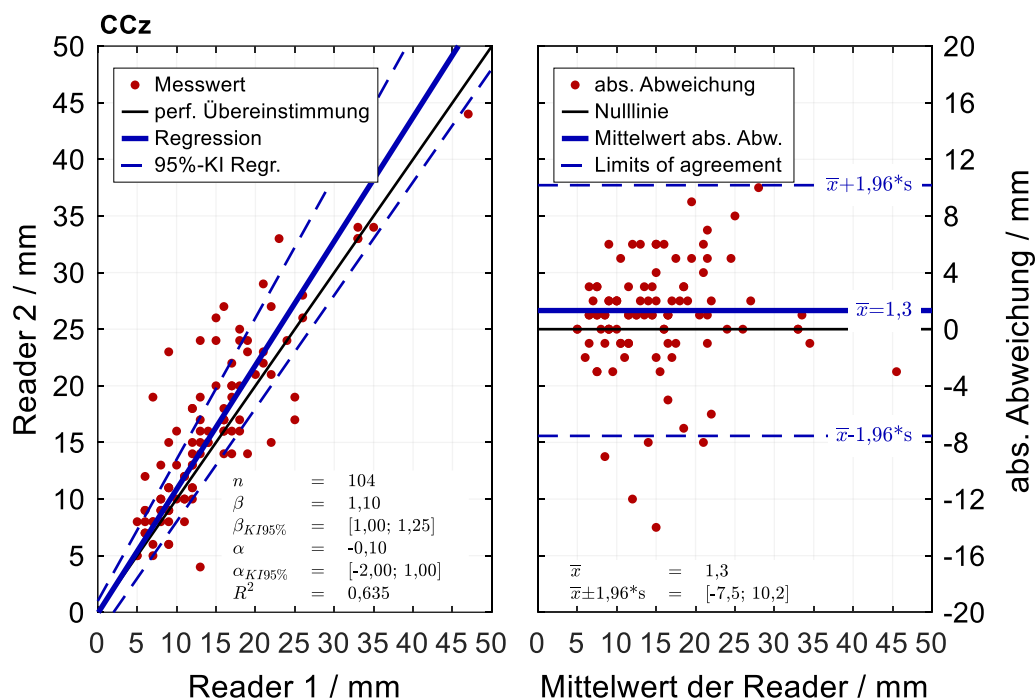


Abbildung 8.4: Inter-Reader Vergleich der Größe des Herdbefundes am Beispiel CCz (MG)

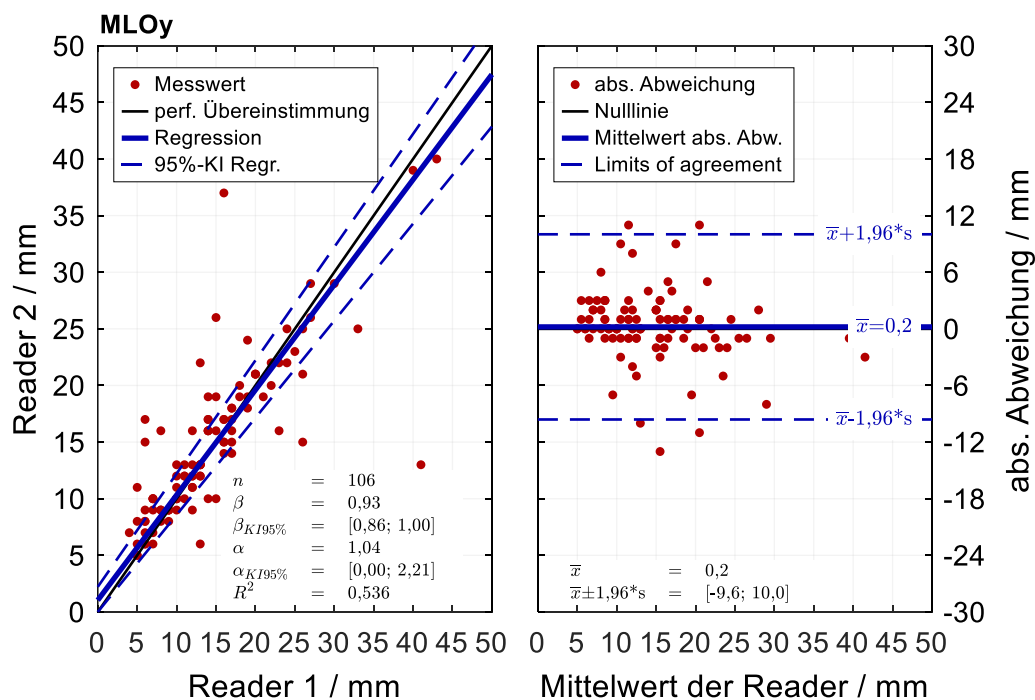


Abbildung 8.5: Inter-Reader-Vergleich der Größe des Herdbefundes am Beispiel MLOy (MG)

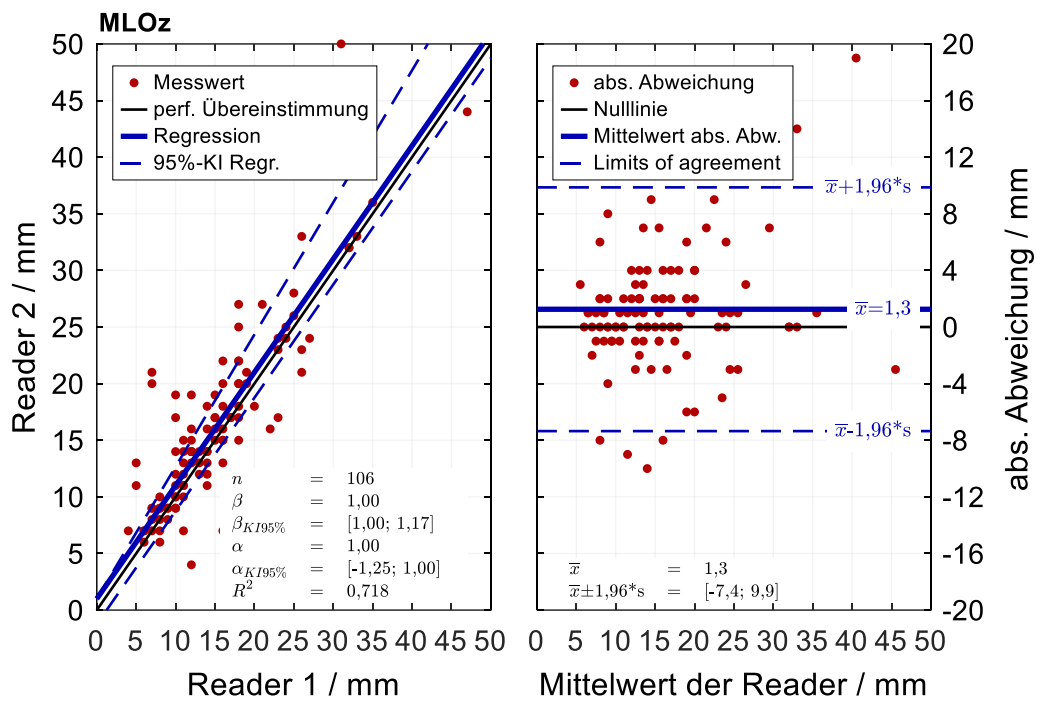


Abbildung 8.6: Inter-Reader-Vergleich der Größe des Herdbefundes am Beispiel MLOz (MG)

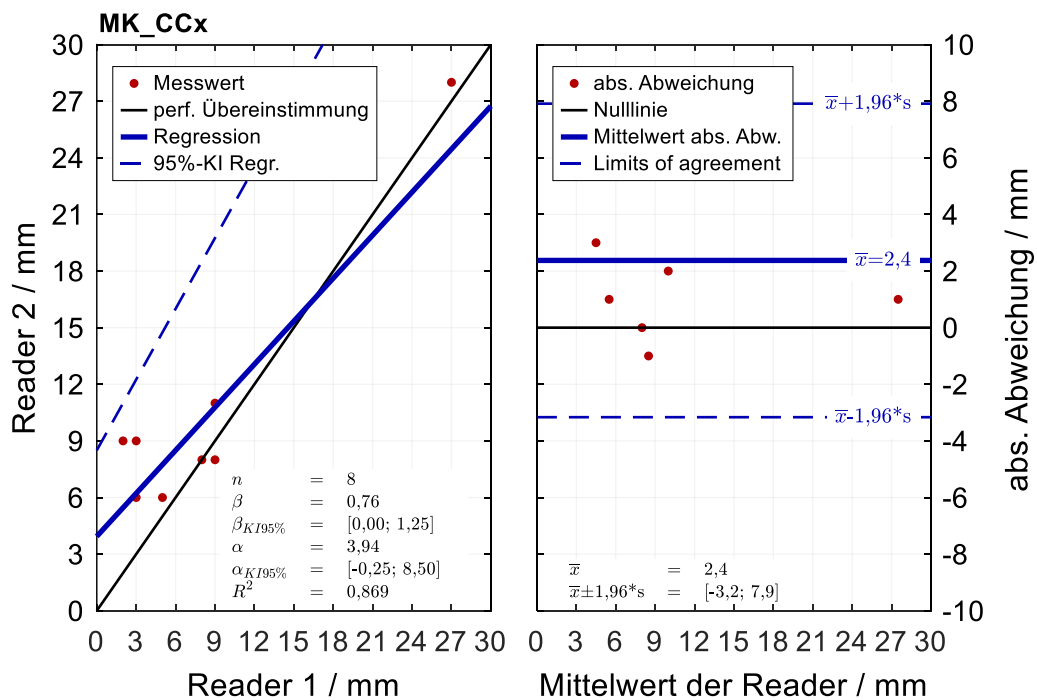


Abbildung 8.7: Inter-Reader-Vergleich der Größe der MK am Beispiel CCx (MG)

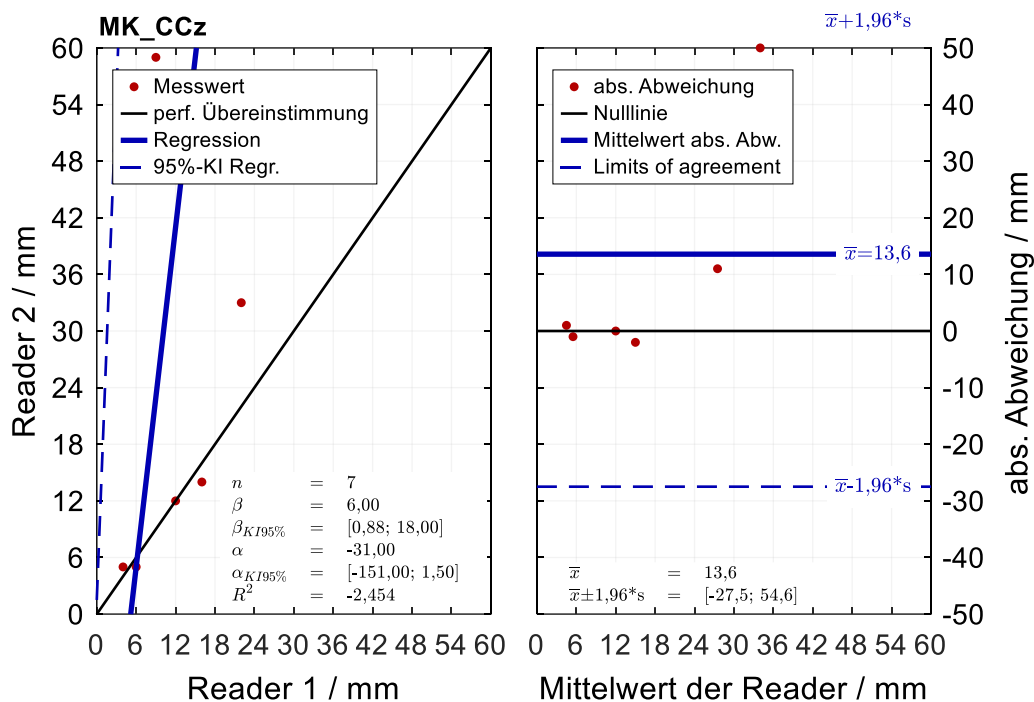


Abbildung 8.8: Inter-Reader-Vergleich der Größe der MK am Beispiel CCy (MG)

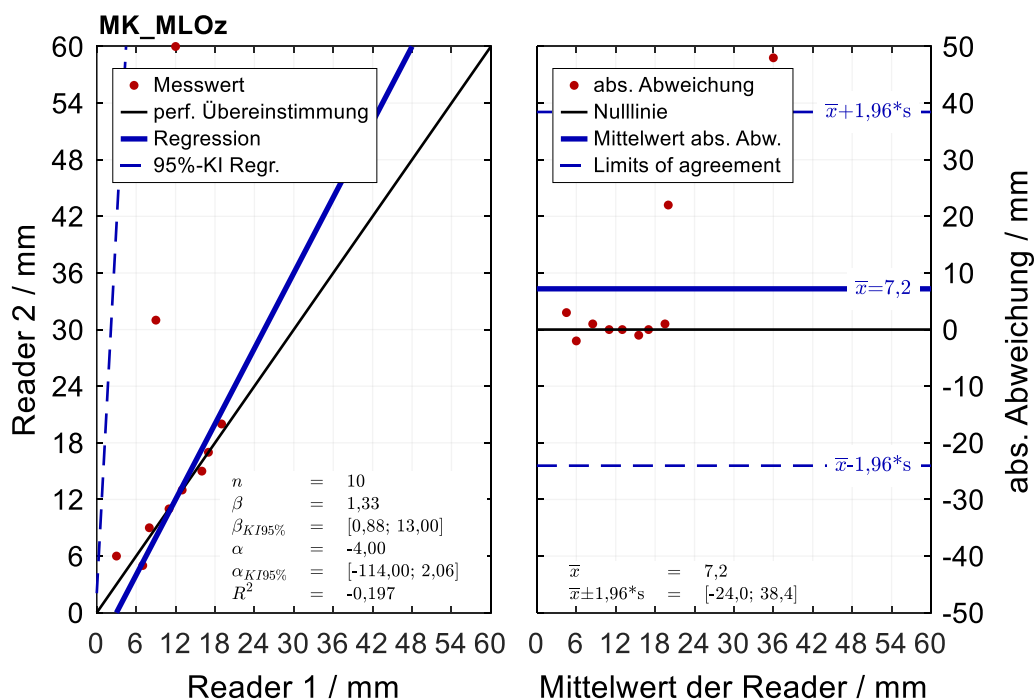


Abbildung 8.9: Inter-Reader-Vergleich der Größe der MK am Beispiel MLOz (MG)

8.5 Weitere Diagramme Inter-Reader-Vergleich PR

8.5.1 Präparatgröße

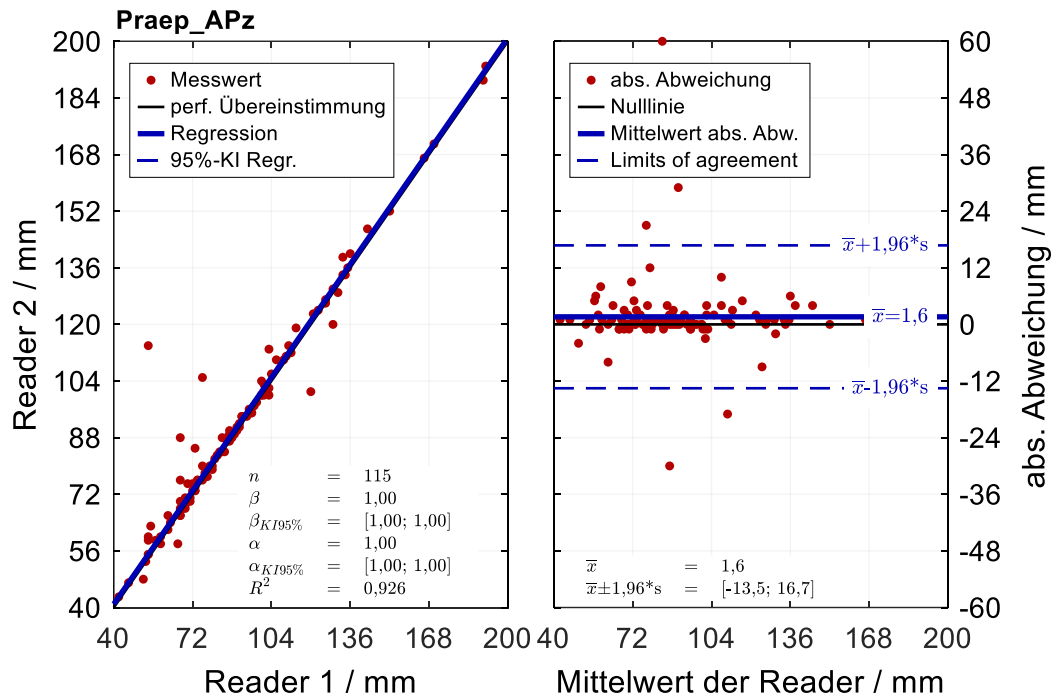


Abbildung 8.10: Inter-Reader-Vergleich der Größe des Präparats am Beispiel APx (PR)

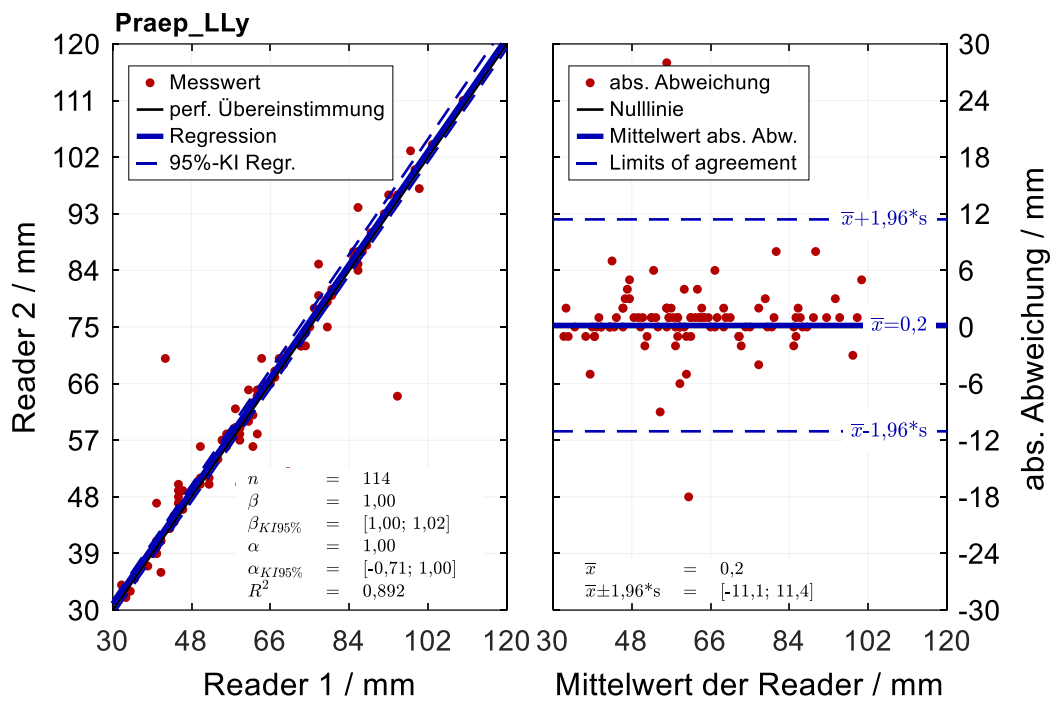


Abbildung 8.11: Inter-Reader-Vergleich der Größe des Präparats am Beispiel LLy (PR)

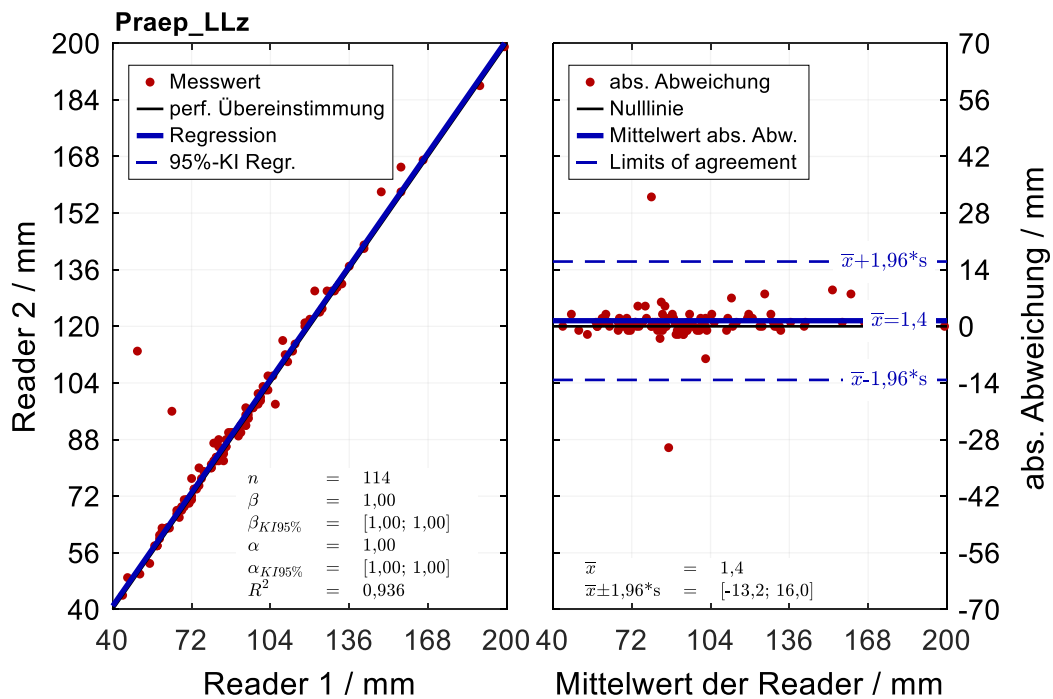


Abbildung 8.12: Inter-Reader-Vergleich der Größe des Präparats am Beispiel LLz (PR)

8.5.2 Größe Herdbefund

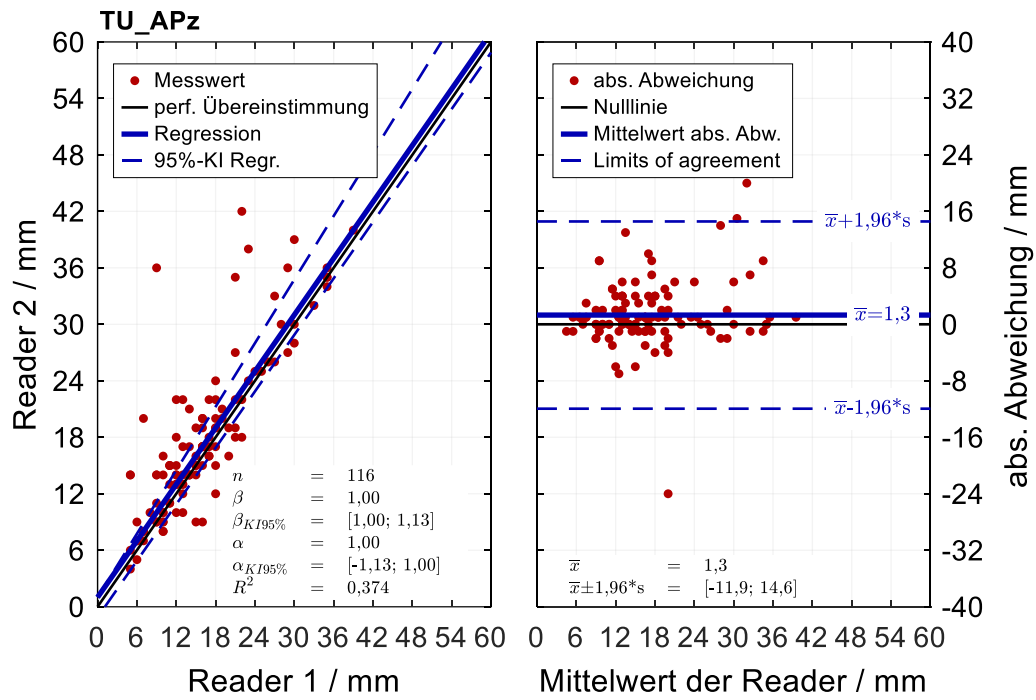


Abbildung 8.13: Inter-Reader-Vergleich der Größe des Herdbefundes am Beispiel APz (PR)

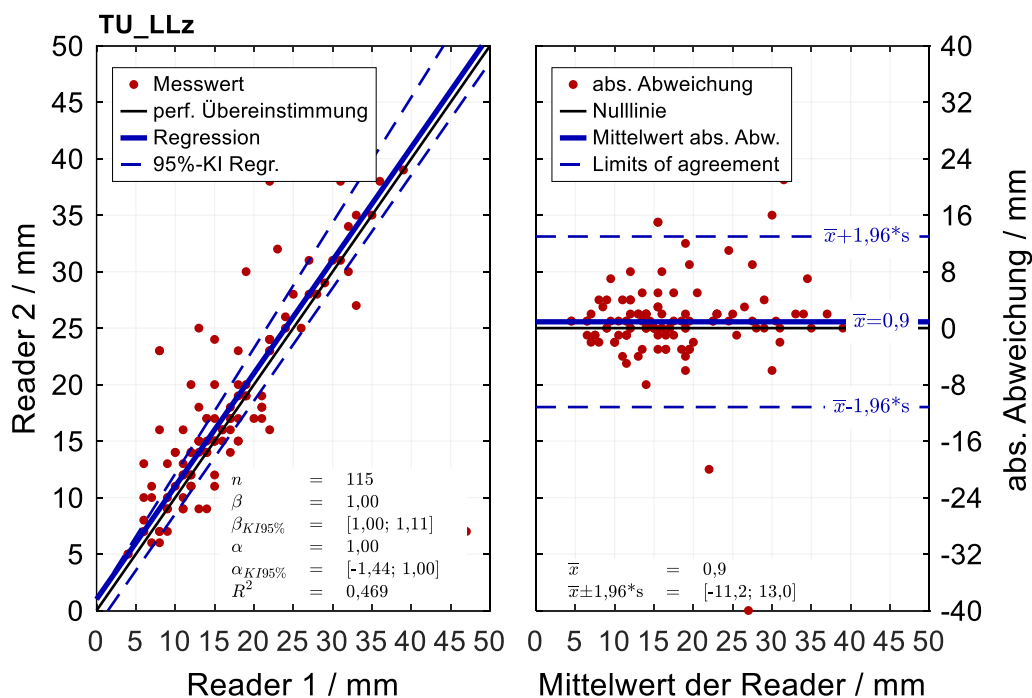


Abbildung 8.14: Inter-Reader-Vergleich der Größe des Herdbefundes am Beispiel LLz (PR)

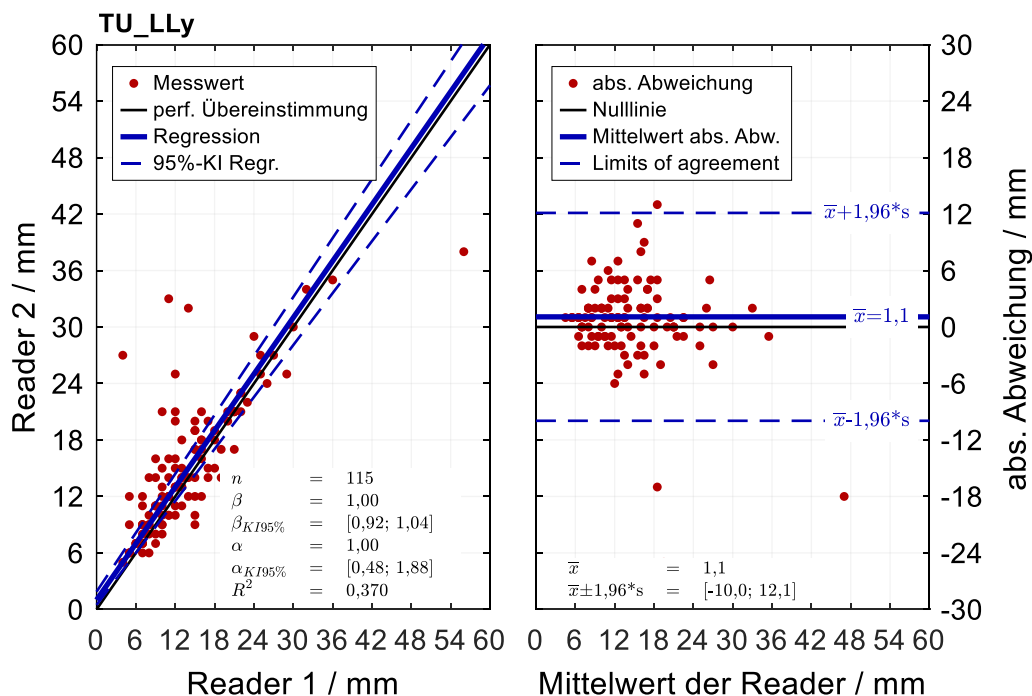


Abbildung 8.15: Inter-Reader-Vergleich der Größe des Herdbefundes am Beispiel LLy (PR)

8.5.3 Randabstand

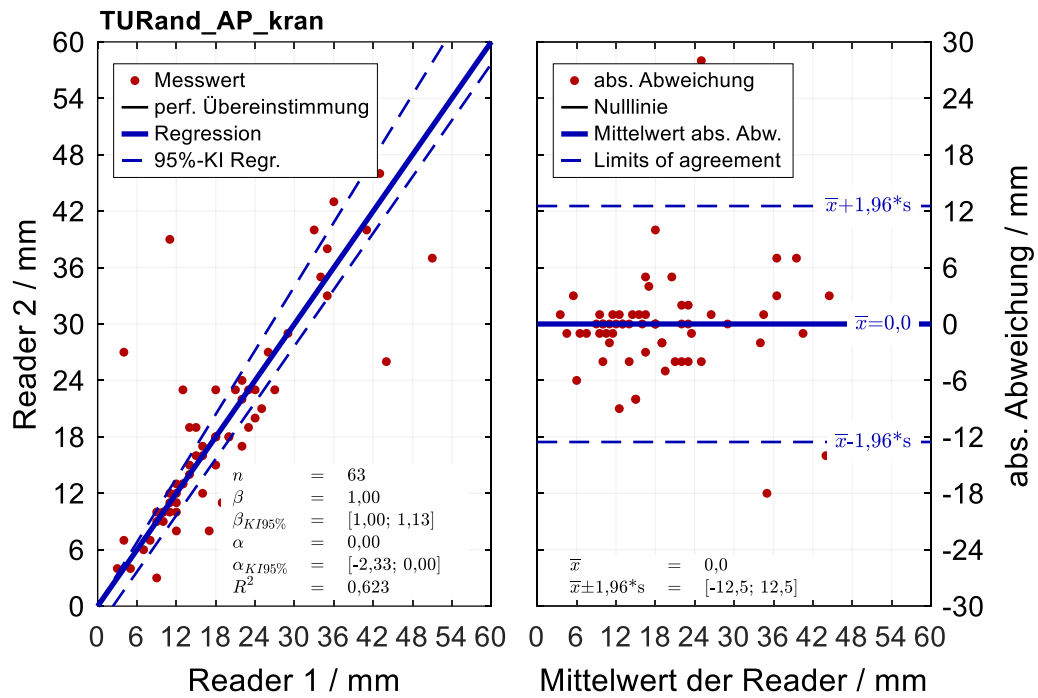


Abbildung 8.16: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel APkranial (PR)

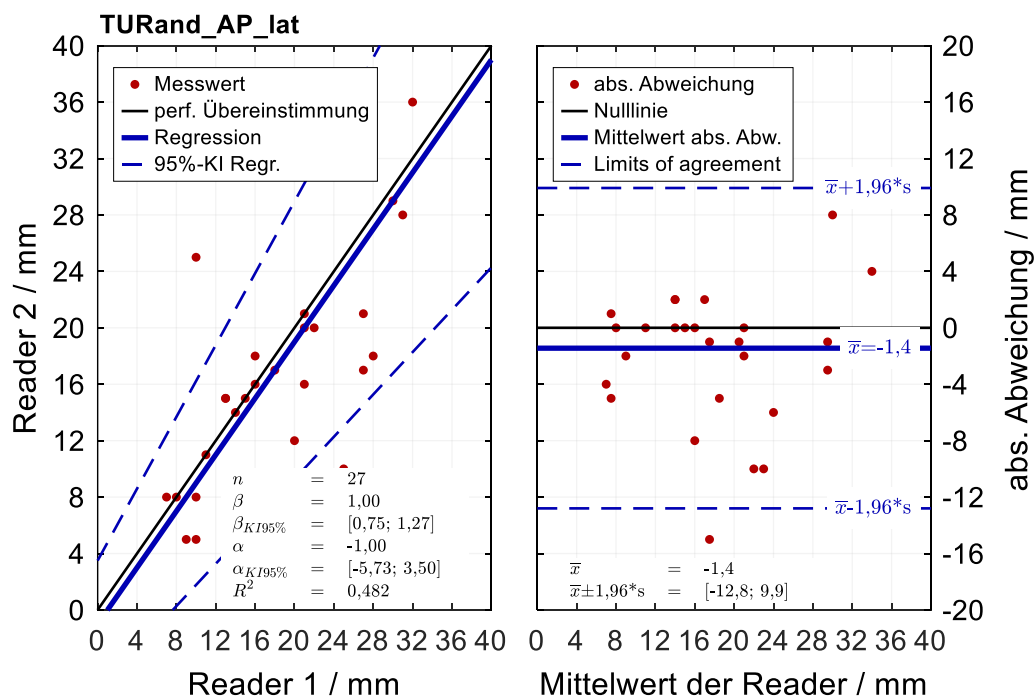


Abbildung 8.17: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel APlateral (PR)

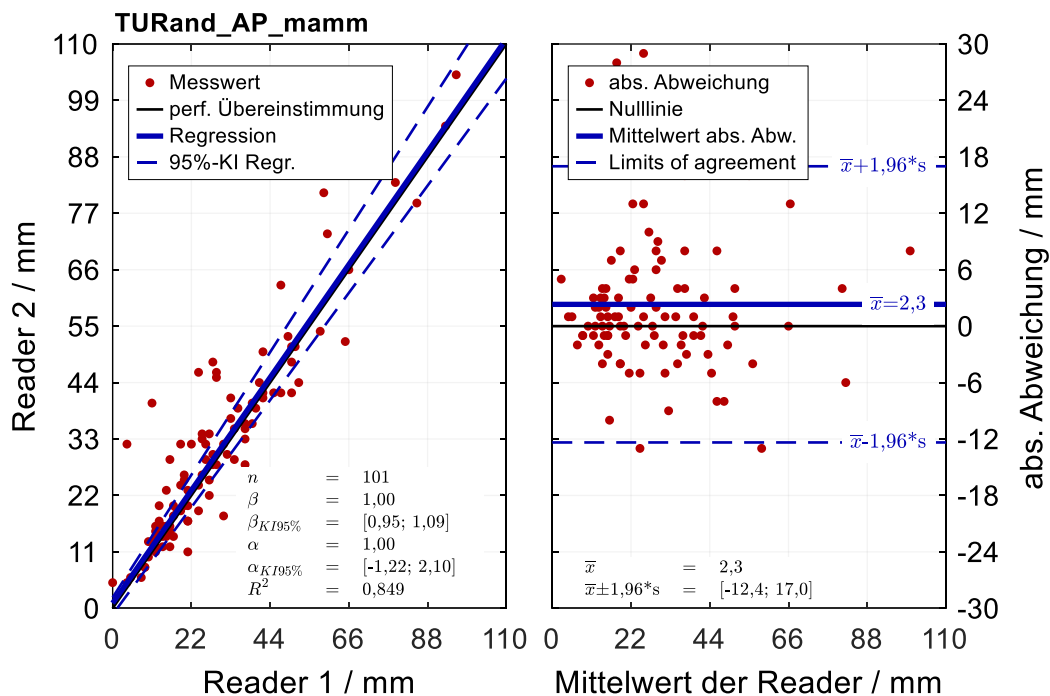


Abbildung 8.18: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel APmammilär (PR)

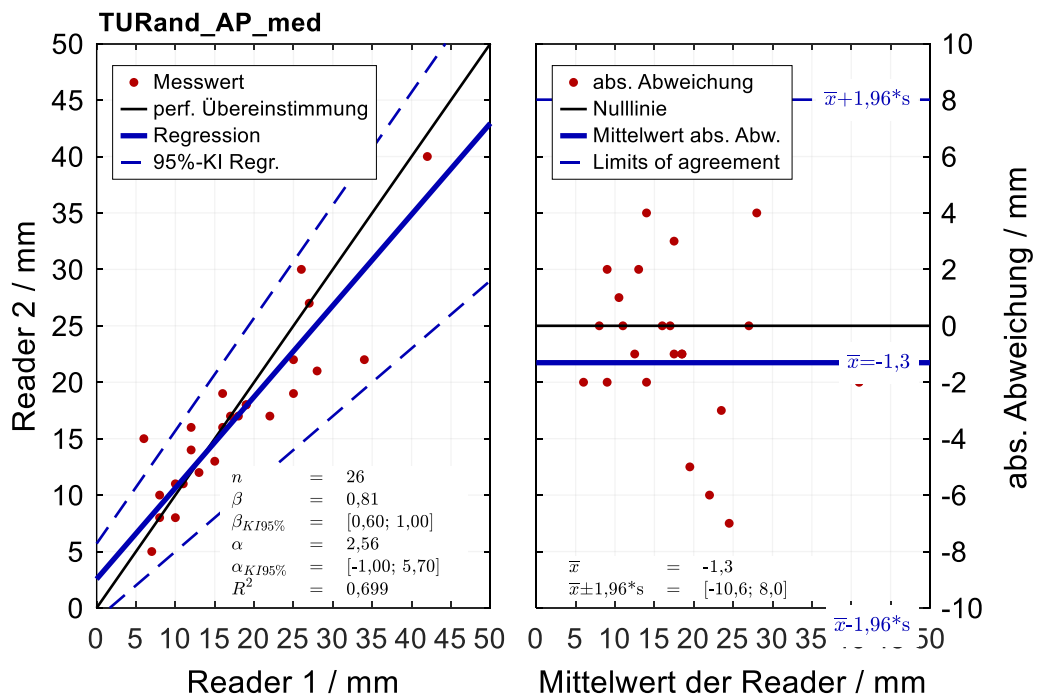


Abbildung 8.19: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel APmedial (PR)

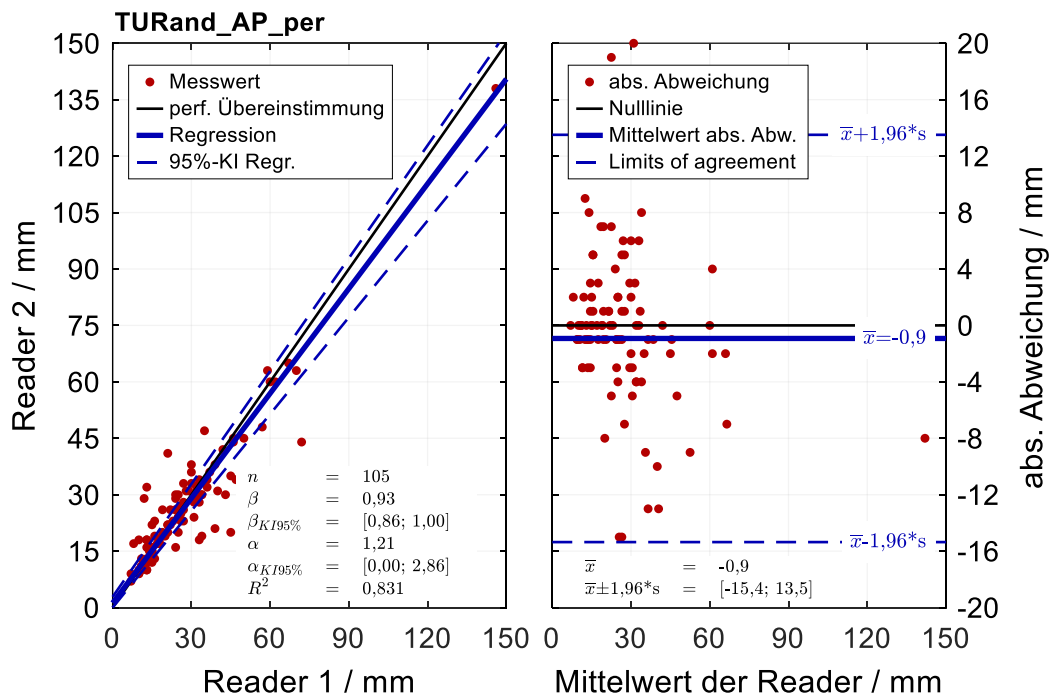


Abbildung 8.20: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel APperipher (PR)

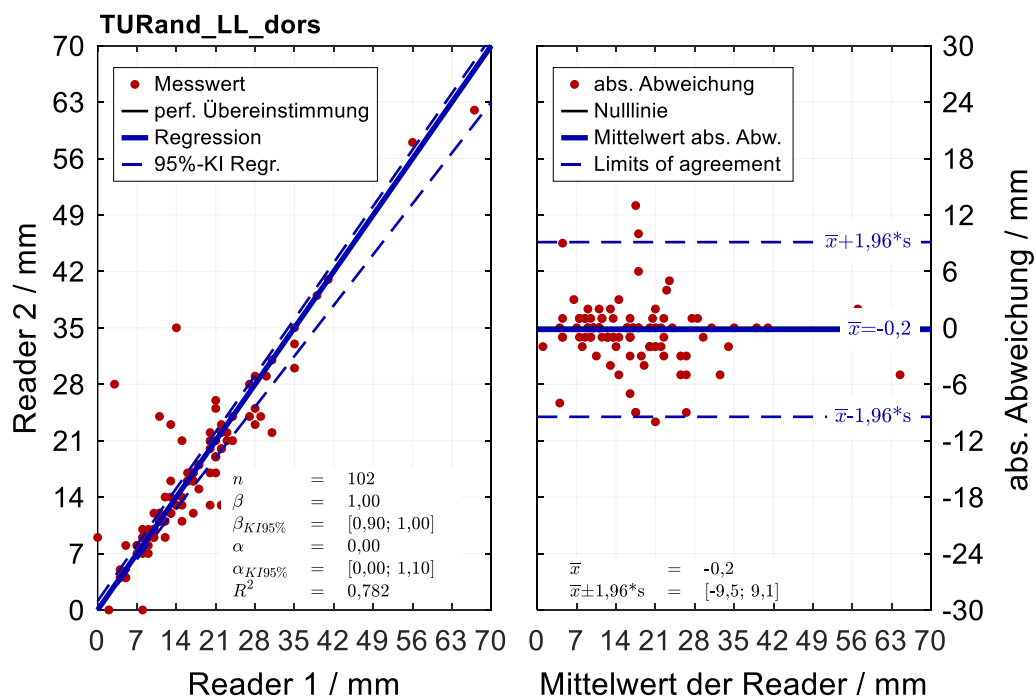


Abbildung 8.21: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel LLdorsal (PR)

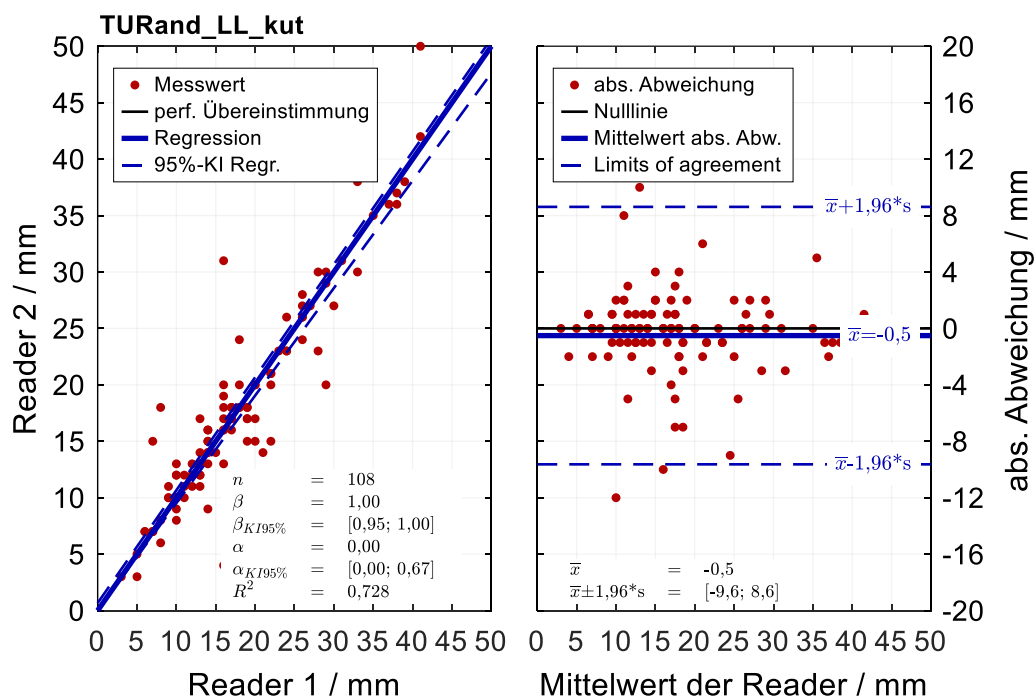


Abbildung 8.22: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel LLkutan (PR)

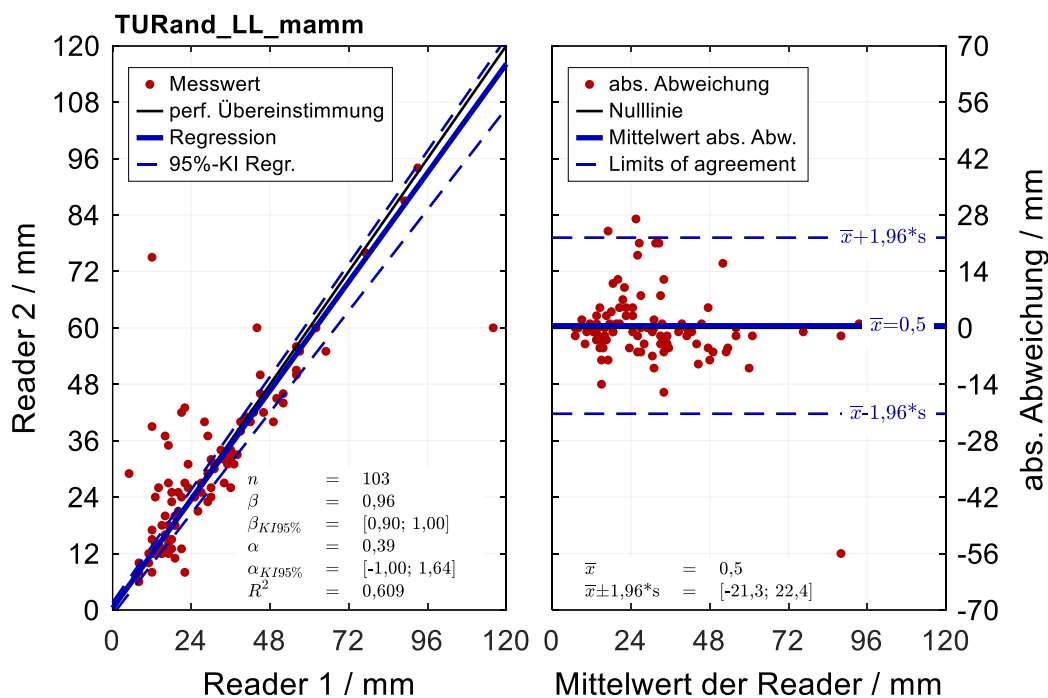


Abbildung 8.23: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel LLmammilär (PR)

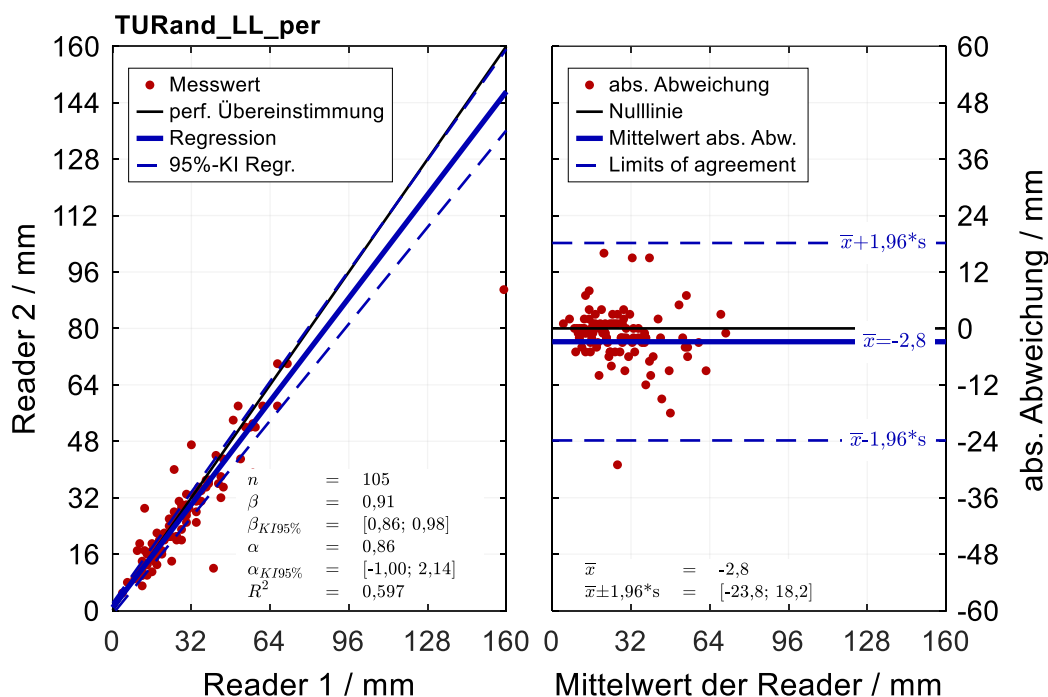


Abbildung 8.24: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel LLperipher (PR)

8.6 Weitere Diagramme Inter-Reader-Vergleich DBT

8.6.1 Präparatgröße

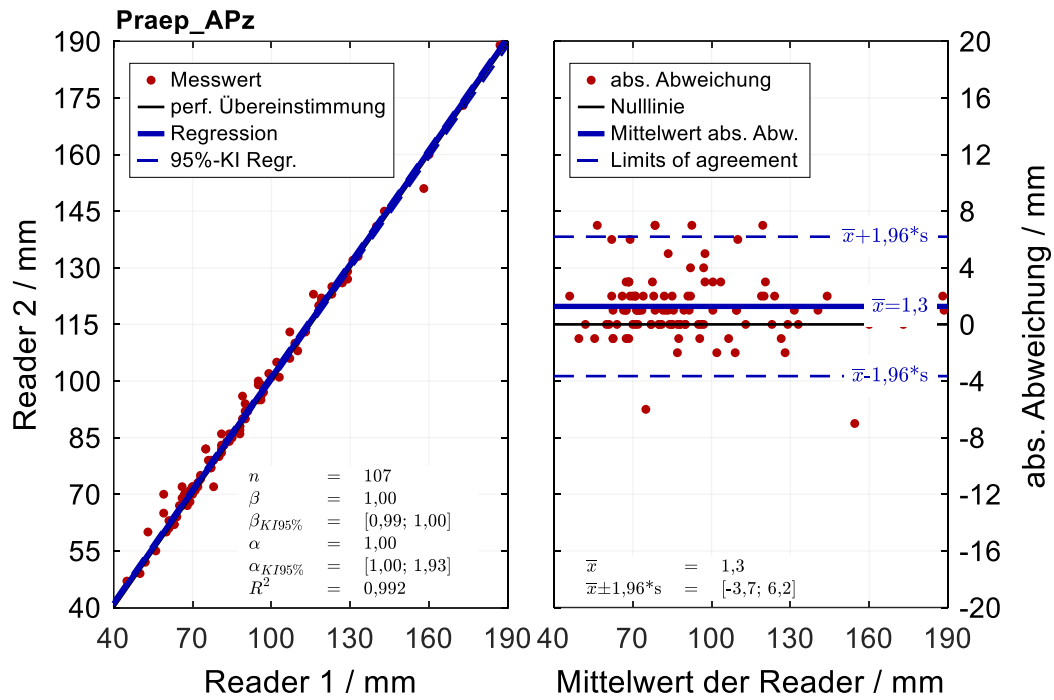


Abbildung 8.25: Inter-Reader-Vergleich der Größe des Präparats am Beispiel APx (DBT)

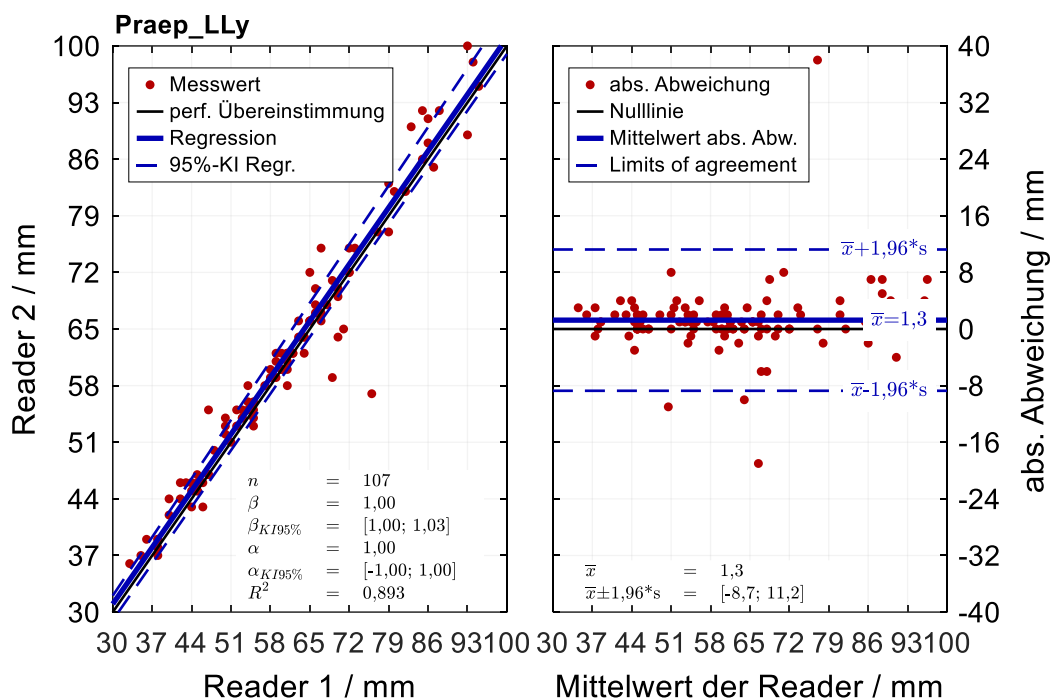


Abbildung 8.26: Inter-Reader-Vergleich der Größe des Präparats am Beispiel LLy (DBT)

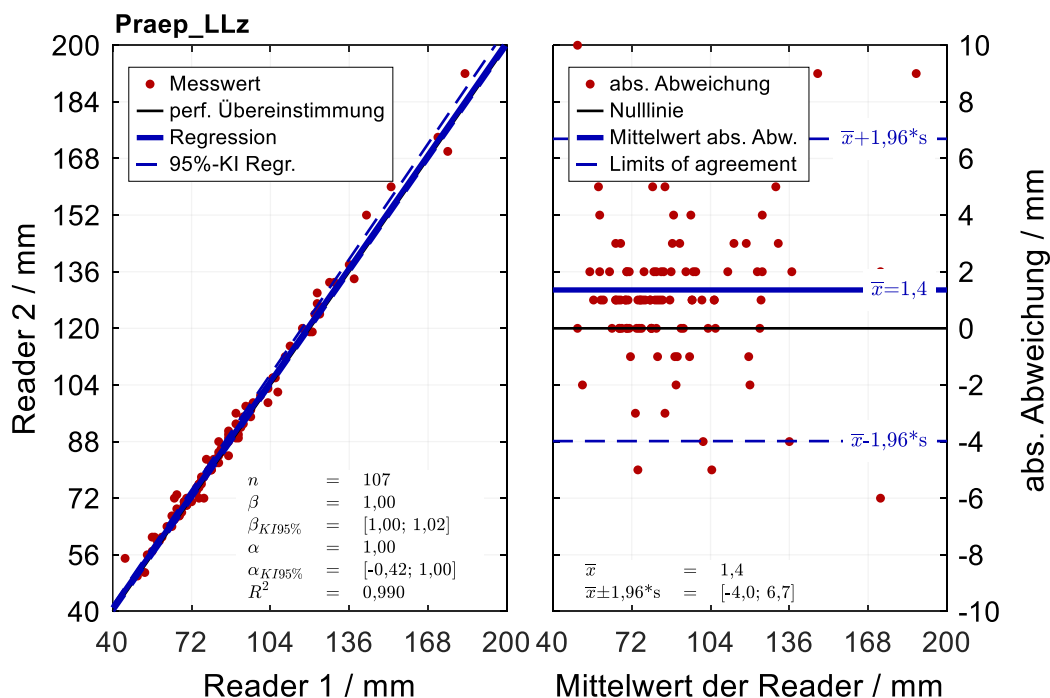


Abbildung 8.27: Inter-Reader-Vergleich der Größe des Präparats am Beispiel LLz (DBT)

8.6.2 Größe Herdbefund

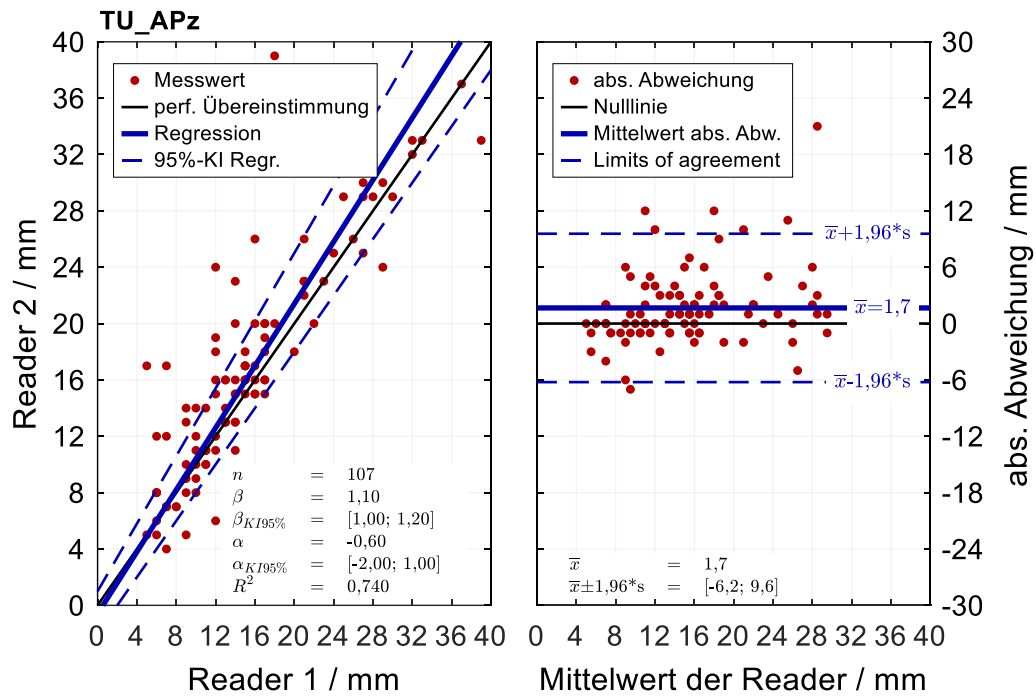


Abbildung 8.28: Inter-Reader-Vergleich der Größe des Herdbefundes am Beispiel APz (DBT)

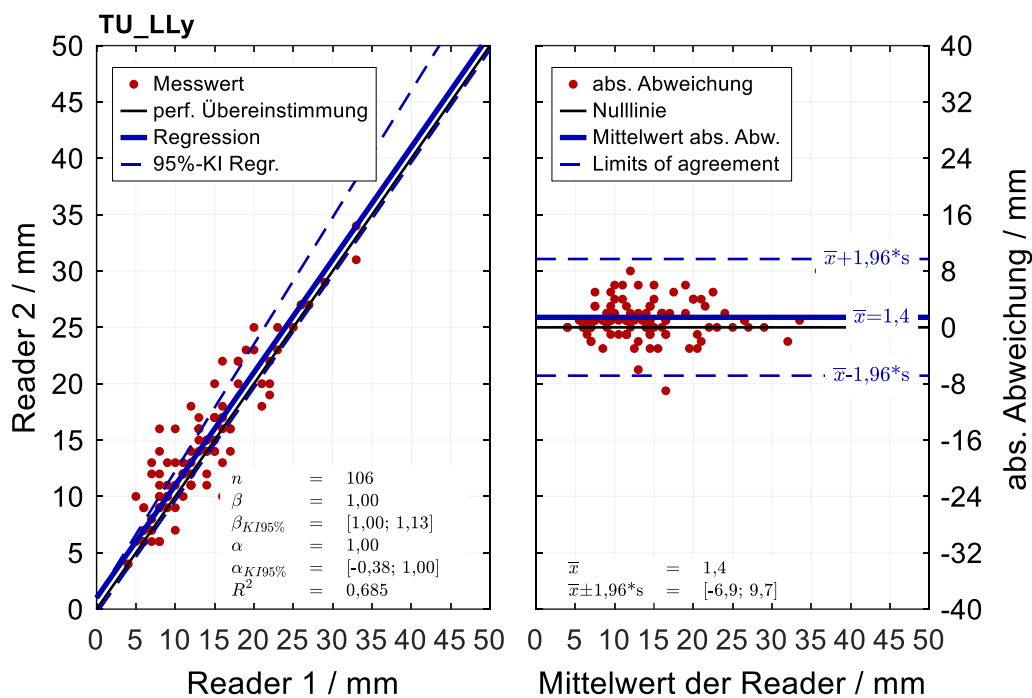


Abbildung 8.29: Inter-Reader-Vergleich der Größe des Herdbefundes am Beispiel LLy (DBT)

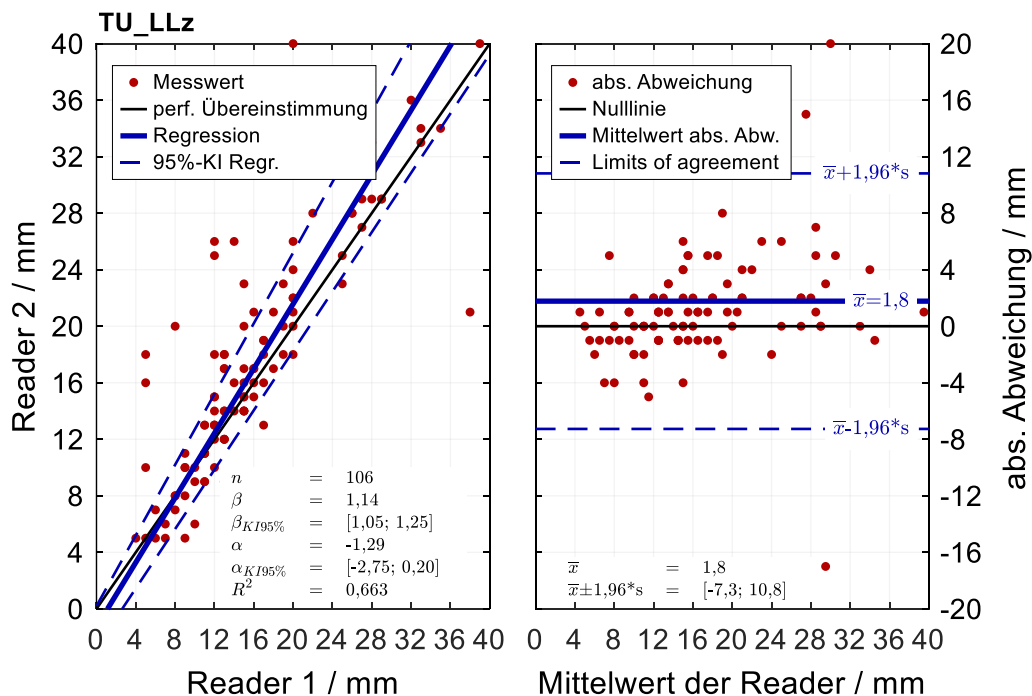


Abbildung 8.30: Inter-Reader-Vergleich der Größe des Herdbefundes am Beispiel LLz (DBT)

8.6.3 Randabstand

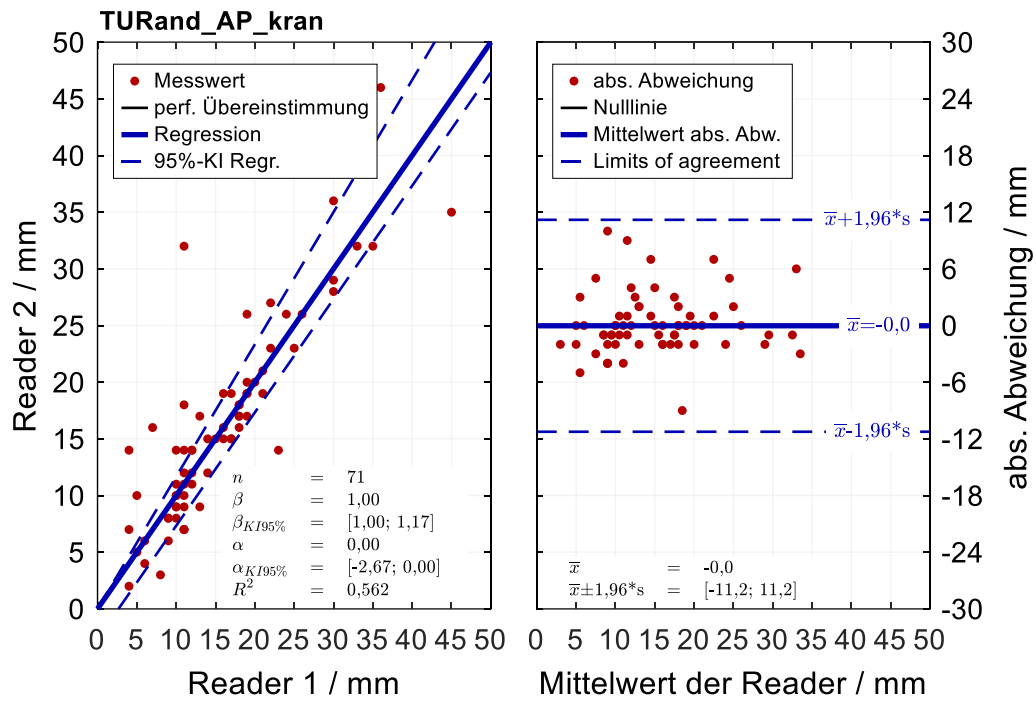


Abbildung 8.31: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel APkranial (DBT)

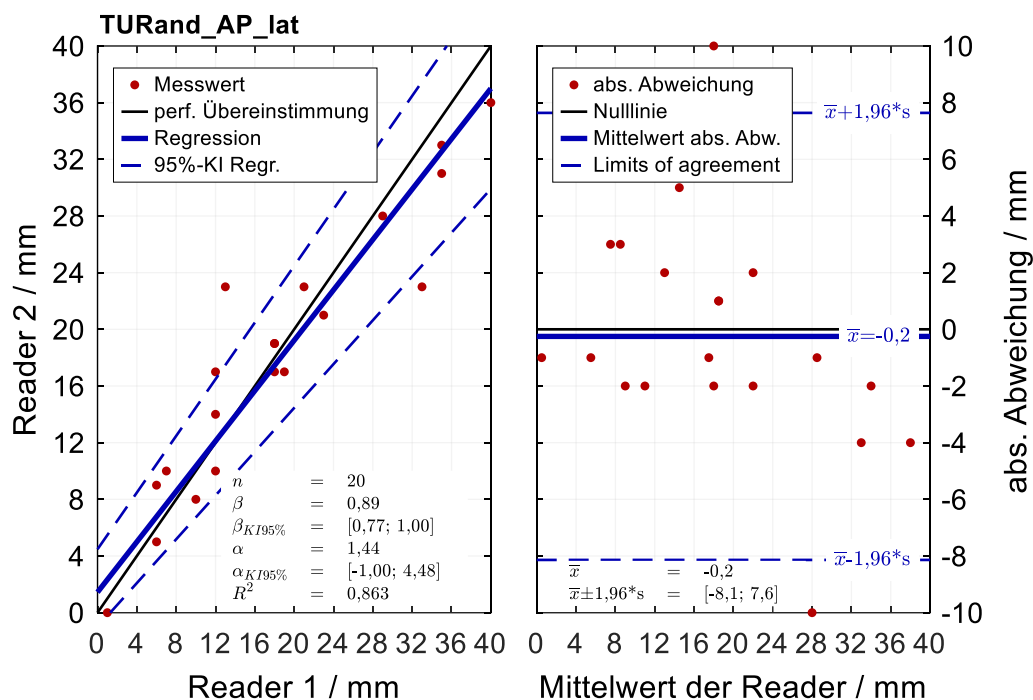


Abbildung 8.32: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel APlateral (DBT)

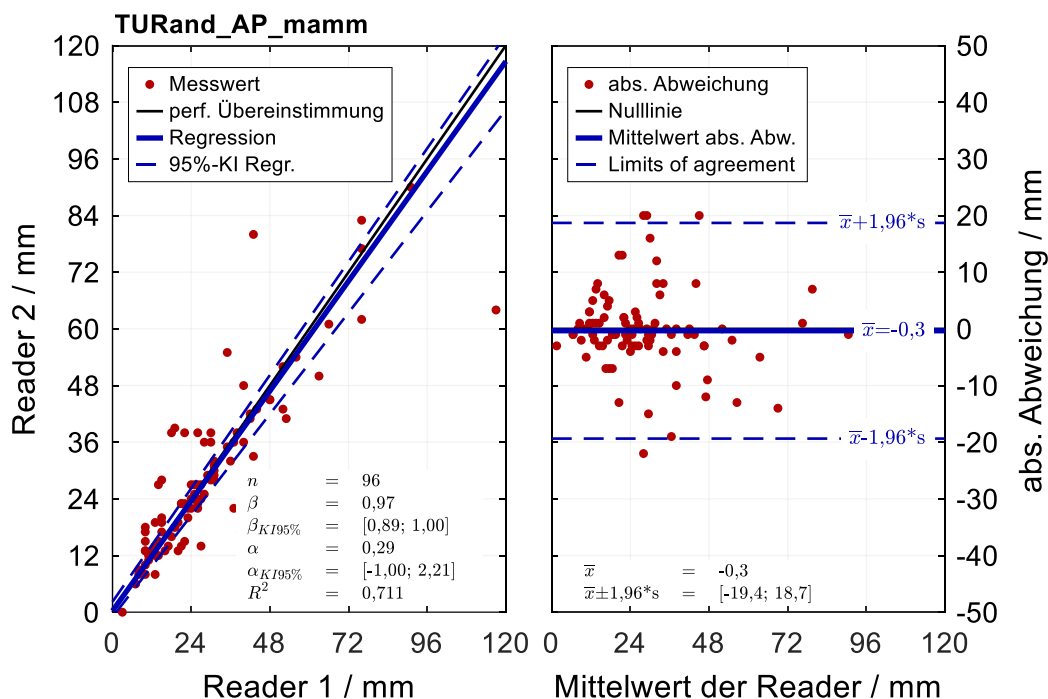


Abbildung 8.33: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel APmammilär (DBT)

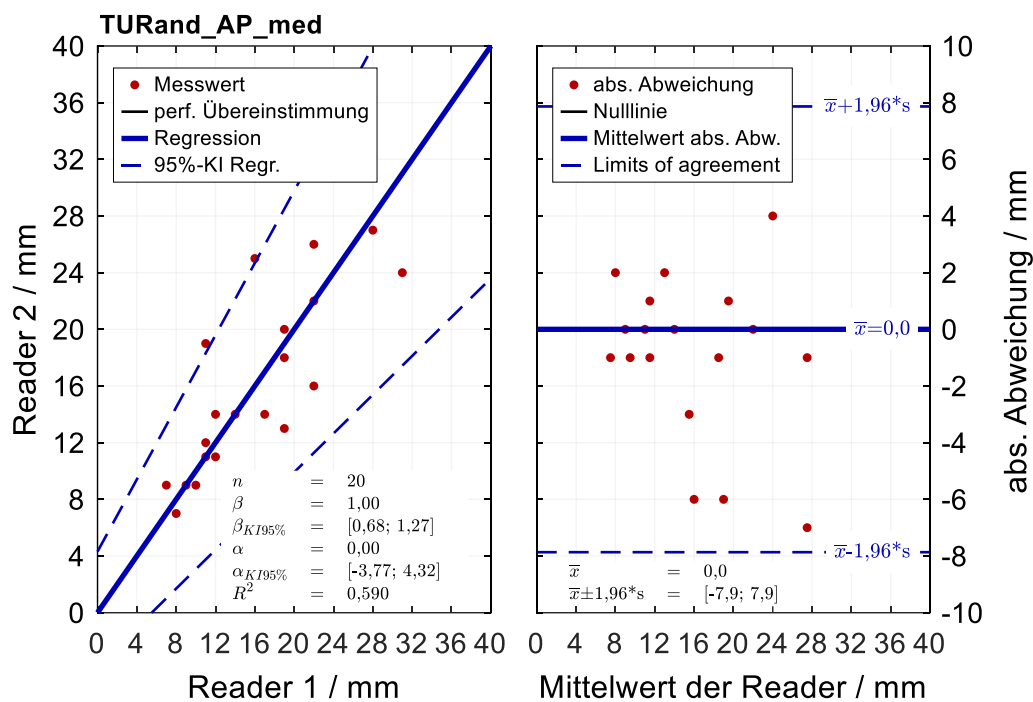


Abbildung 8.34: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel APmedial (DBT)

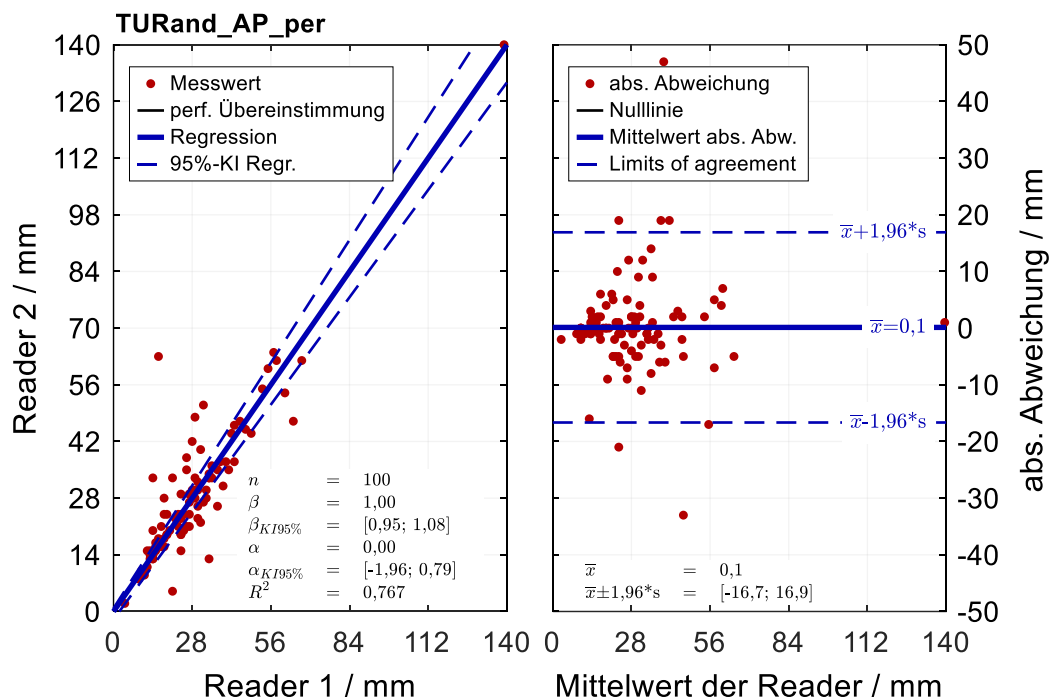


Abbildung 8.35: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel APperipher (DBT)

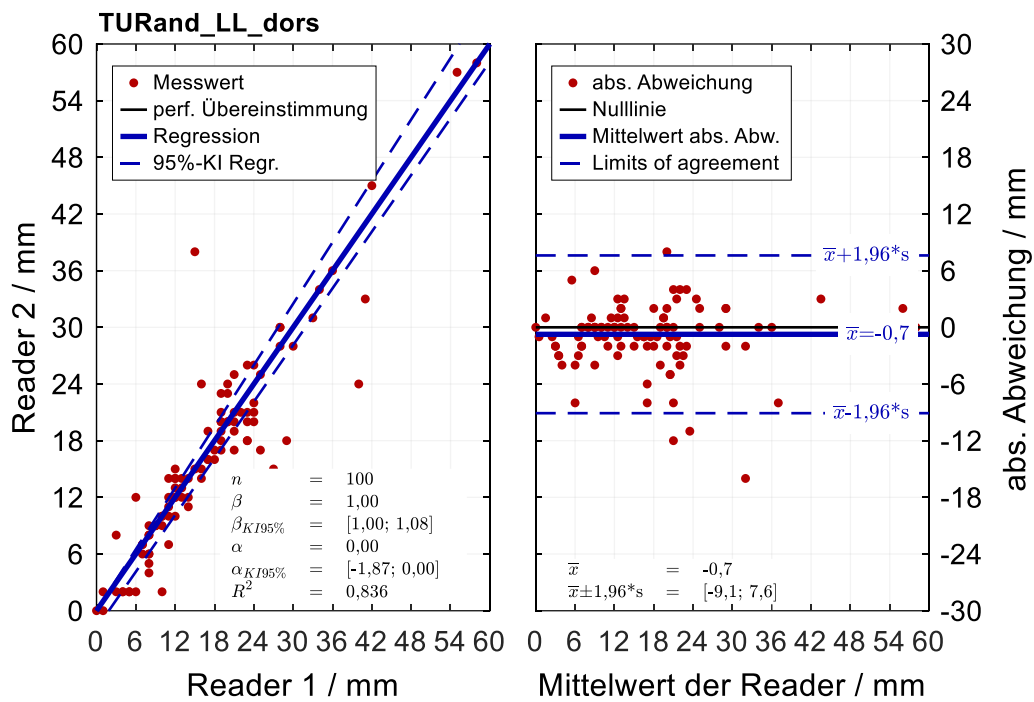


Abbildung 8.36: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel LLdorsal (DBT)

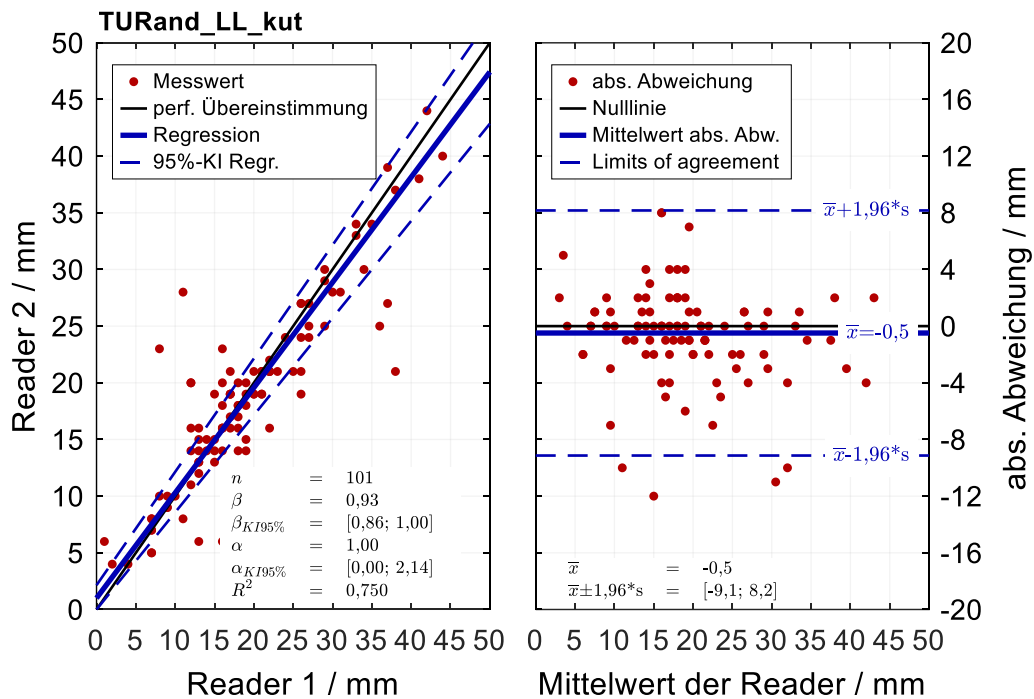


Abbildung 8.37: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel LLkutan (DBT)

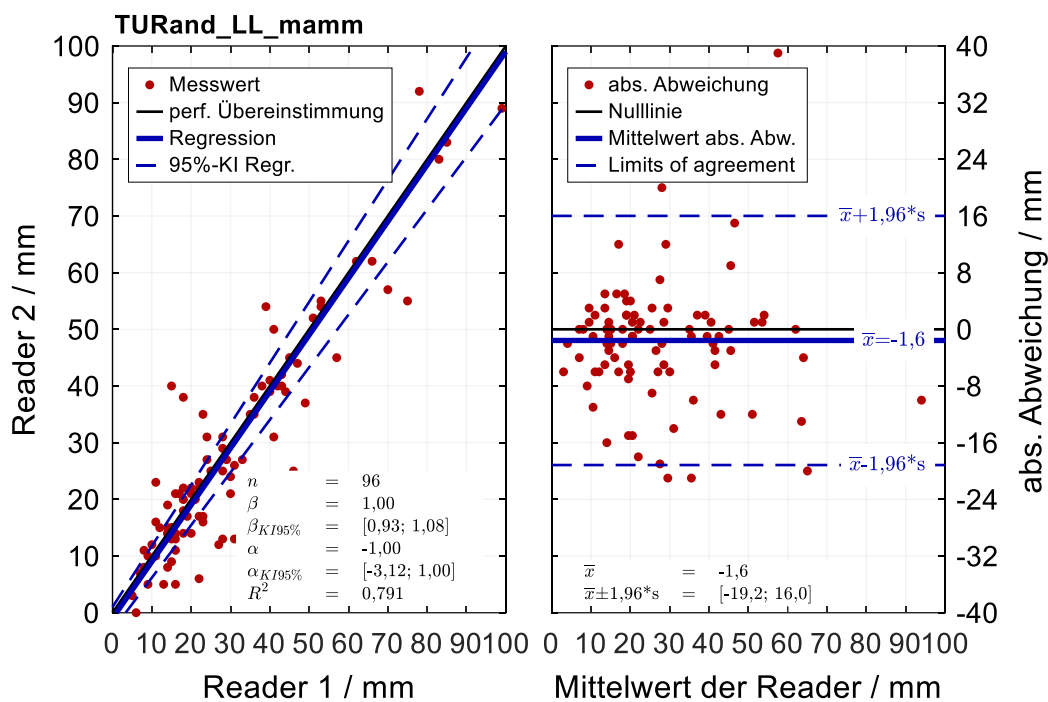


Abbildung 8.38: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel LLmammilär (DBT)

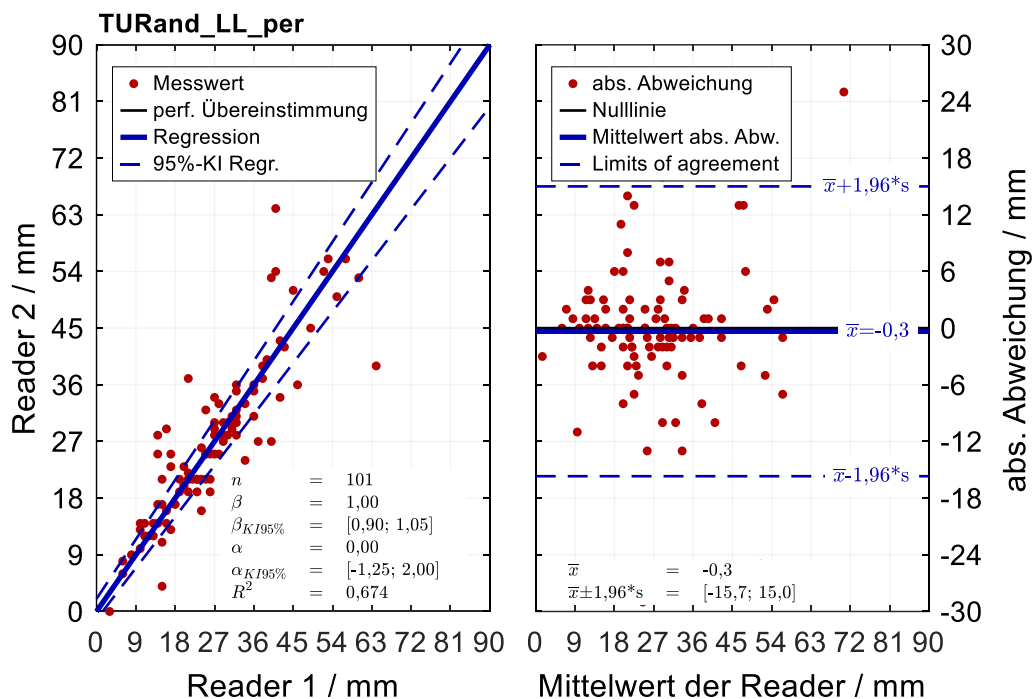


Abbildung 8.39: Inter-Reader-Vergleich der Größe des Randabstandes am Beispiel LLperipher (DBT)

8.7 Vergleich zum Goldstandard – Weitere Diagramme und Kontingenztafeln

8.7.1 Größe Herdbefund (Reader 2)

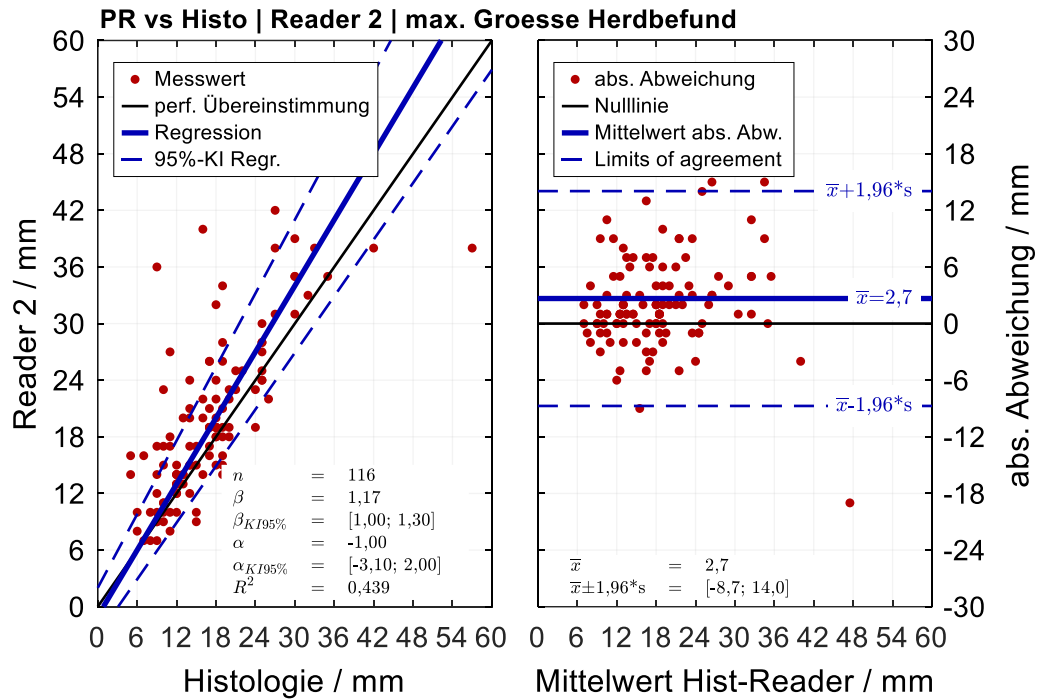


Abbildung 8.40: Vergleich der maximalen Größe des Herdbefundes zwischen PR und Goldstandard (Reader 2)

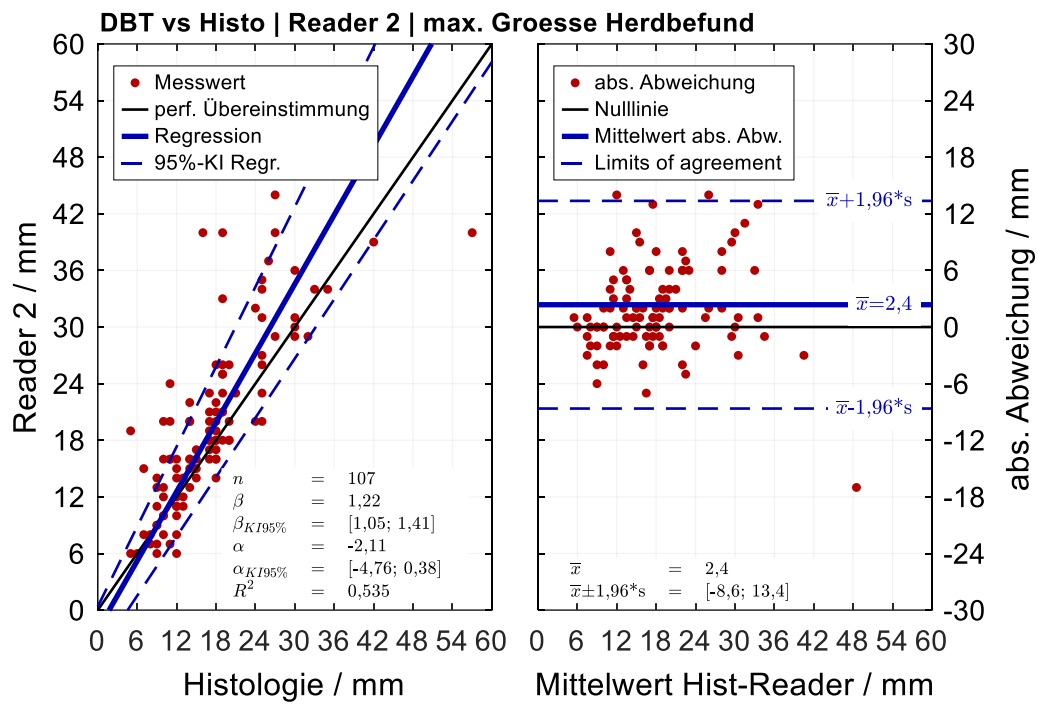


Abbildung 8.41: Vergleich der maximalen Größe des Herdbefundes zwischen DBT und Goldstandard (Reader 2)

8.7.2 Randabstand (Reader 2)

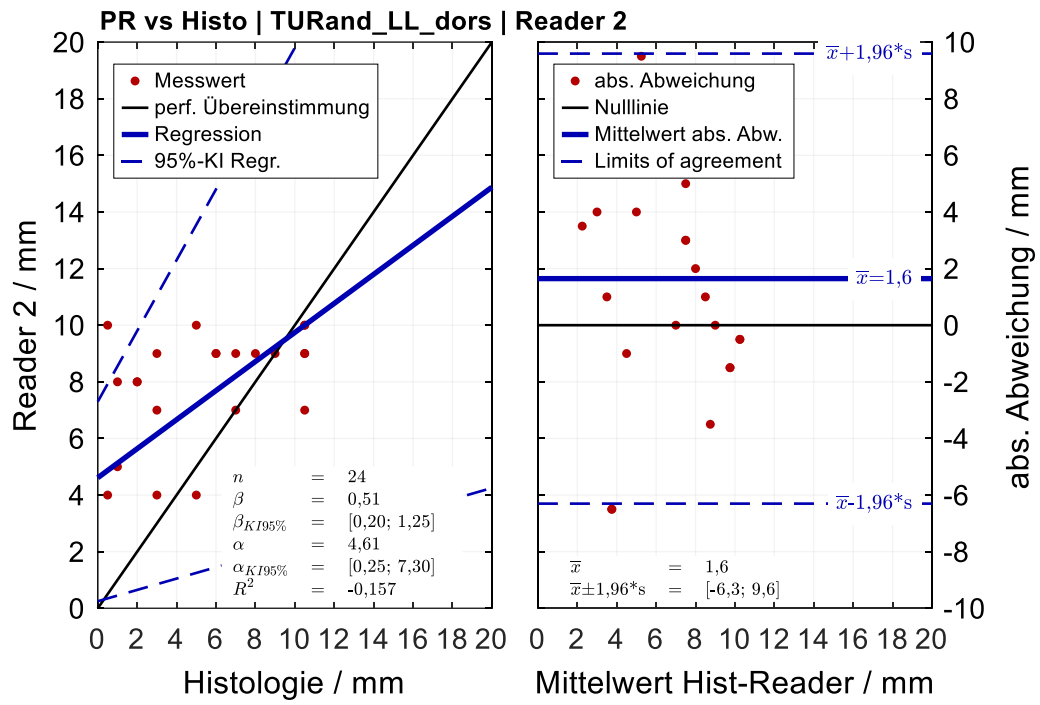


Abbildung 8.42: Vergleich des Randabstandes LLdorsal zwischen PR und Goldstandard (Reader 2)

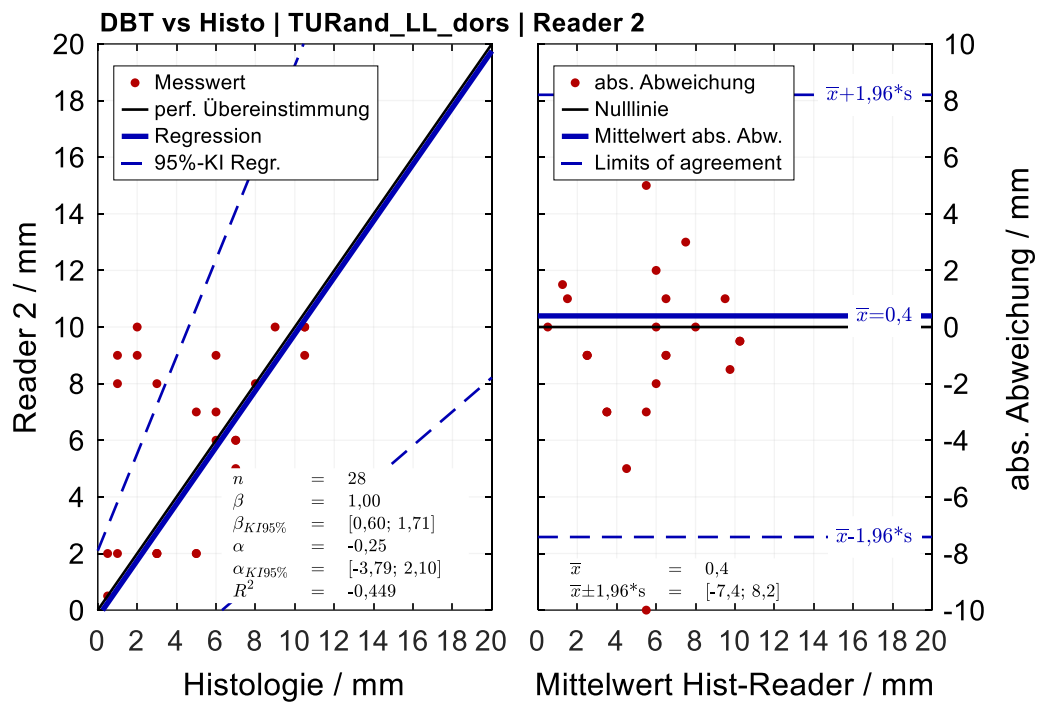


Abbildung 8.43: Vergleich des Randabstandes LLdorsal zwischen DBT und Goldstandard (Reader 2)

8.7.3 Kontingenztafeln R0/R1-Vergleich (Reader 2)

Tabelle 8.1: Kontingenztafel R0/R1-Vergleich PR und Histologie (Reader 2)

Kontingenztafel intraoperative PR vs Histo Reader 2 R0/R1 (abs.)

		PR Reader 2		Rand- häufigkeit		
		R0	R1			
Histo	R0	13	2	15	Spezifität	0,867
	R1	48	53	101	Sensitivität	0,525
Rand- häufigkeit		61	55	116		

Kontingenztafel intraoperative PR vs Histo Reader 2 R0/R1 (rel.)

		PR Reader 2		Rand- häufigkeit		
		R0	R1			
Histo	R0	0,11207	0,017241	0,12931		
	R1	0,41379	0,4569	0,87069		
Rand- häufigkeit		0,52586	0,47414			

p0 0,56897

pe 0,48083

K2 0,16977

unterer
Wert -0,00381oberer
Wert 0,34335

Tabelle 8.2: Kontingenztafel R0/R1-Vergleich PR und Histologie (Reader 2, Datensatz 2)

Kontingenztafel intraoperative PR vs Histo Reader 2 R0/R1 (abs.)

		PR Reader 2		Rand- häufigkeit		
		R0	R1			
Histo	R0	13	0	13	Spezifität	1,000
	R1	6	13	19	Sensitivität	0,684
Rand- häufigkeit		19	13	32		

Kontingenztafel intraoperative PR vs Histo Reader 2 R0/R1 (rel.)

		PR Reader 2		Rand- häufigkeit
		R0	R1	
Histo	R0	0,40625	0	0,40625
	R1	0,1875	0,40625	0,59375
Rand- häufigkeit		0,59375	0,40625	

p0 0,8125

pe 0,48242

K2 0,63774

unterer Wert 0,37645 **oberer Wert** 0,89902

Tabelle 8.3: Kontingenztabelle R0/R1-Vergleich DBT und Histologie (Reader 2)

Kontingenztabelle intraoperative DBT vs Histo Reader 2 R0/R1 (abs.)

		DBT Reader 2		Rand- häufigkeit		
		R0	R1			
Histo	R0	10	2	12	Spezifität	0,833
	R1	39	56	95	Sensitivität	0,589
Rand- häufigkeit		49	58	107		

Kontingenztabelle intraoperative DBT vs Histo Reader 2 R0/R1 (rel.)

		DBT Reader 2		Rand- häufigkeit
		R0	R1	
Histo	R0	0,093458	0,018692	0,11215
	R1	0,36449	0,52336	0,88785
Rand- häufigkeit		0,45794	0,54206	

p0 0,61682

pe 0,53262

K2 0,18015

unterer
Wert -0,016939oberer
Wert 0,37725

Tabelle 8.4: Kontingenztabelle R0/R1-Vergleich DBT und Histologie (Reader 2, Datensatz 2)

Kontingenztabelle intraoperative DBT vs Histo Reader 2 R0/R1 (abs.)

		DBT Reader 2		Rand- häufigkeit		
		R0	R1			
Histo	R0	10	0	10	Spezifität	1,000
	R1	7	12	19	Sensitivität	0,632
Rand- häufigkeit		17	12	29		

Kontingenztabelle intraoperative DBT vs Histo Reader 2 R0/R1 (rel.)

		DBT Reader 2		Rand- häufigkeit		
		R0	R1			
Histo	R0	0,34483	0	0,34483		
	R1	0,24138	0,41379	0,65517		
Rand- häufigkeit		0,58621	0,41379			

p0 0,75862

pe 0,47325

K2 0,54176

**unterer
Wert** 0,24609

**oberer
Wert** 0,83743

8.7.4 Sensitivität des R0/R1-Vergleichs auf Randabstandskriterien (Reader 2)

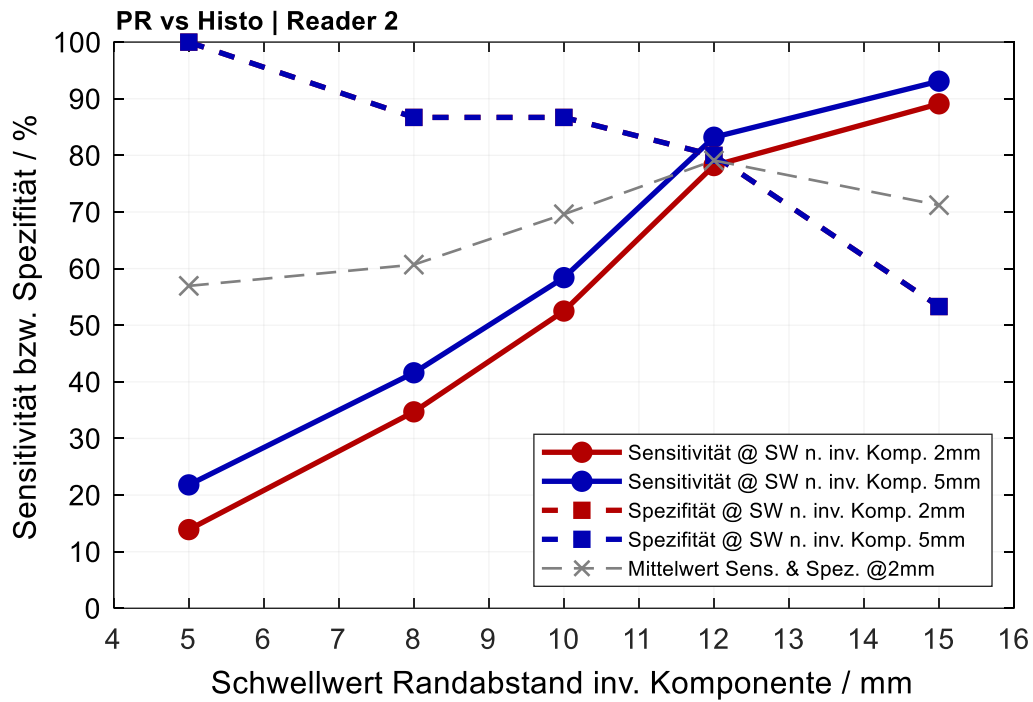


Abbildung 8.44: Sensitivitätsanalyse der R0/R1-Diagnose der PR auf Randabstandskriterien (Reader 2)

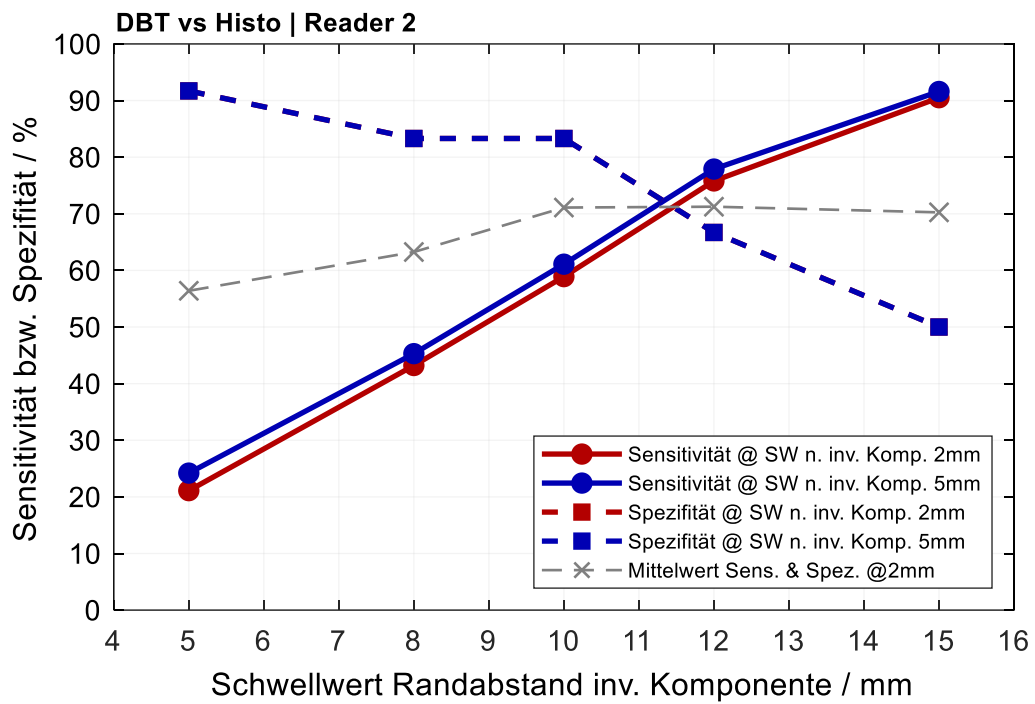


Abbildung 8.45: Sensitivitätsanalyse der R0/R1-Diagnose der DBT auf Randabstandskriterien (Reader 2)

8.7.5 Sensitivität des R0/R1-Vergleichs auf Randabstandskriterien für originale Diagnosen der Histologie (Reader 1 und 2)

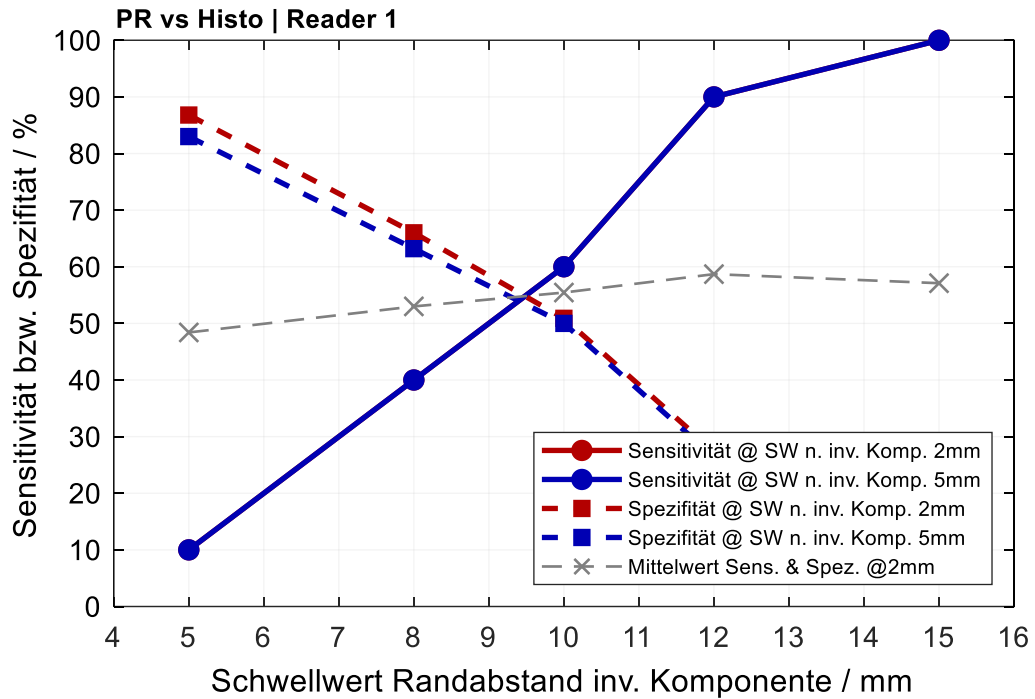


Abbildung 8.46: Sensitivitätsanalyse der R0/R1-Diagnose der PR auf Randabstandskriterien (Reader 1/ Originale R0/R1-Diagnose)

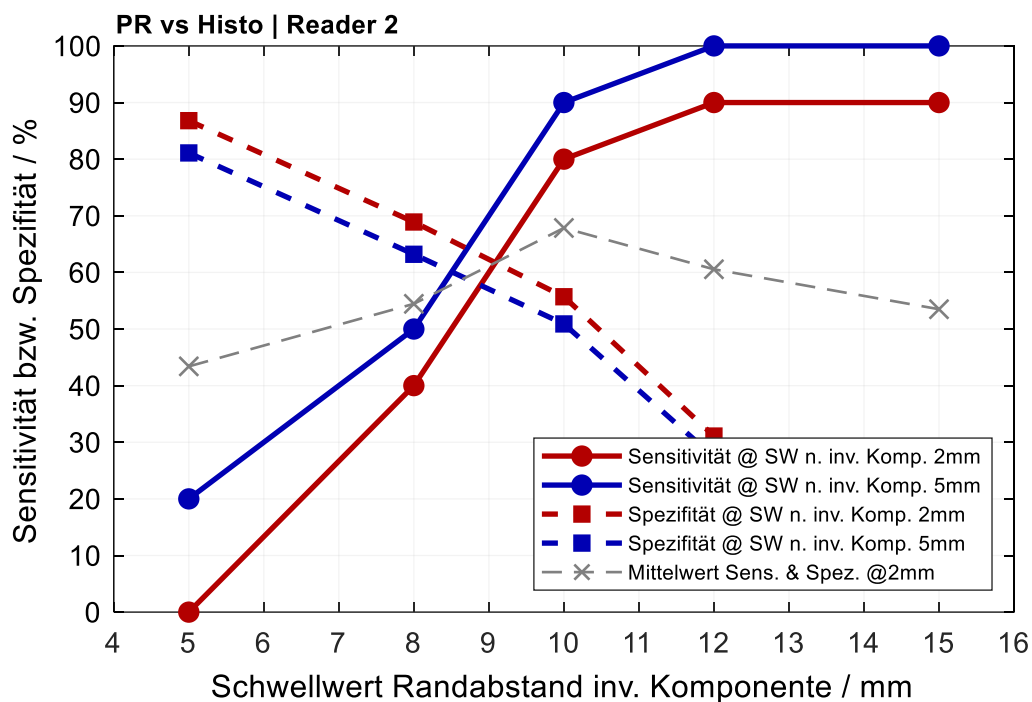


Abbildung 8.47: Sensitivitätsanalyse der R0/R1-Diagnose der PR auf Randabstandskriterien (Reader 2/ Originale R0/R1-Diagnose)

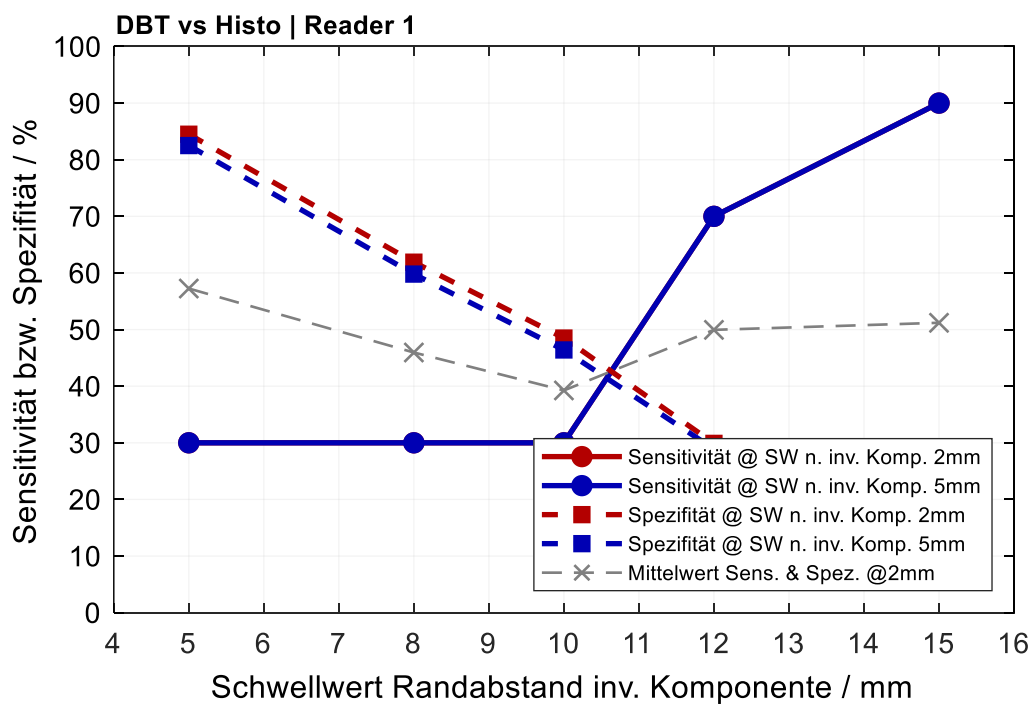


Abbildung 8.48: Sensitivitätsanalyse der R0/R1-Diagnose der DBT auf Randabstandskriterien (Reader 1/ Originale R0/R1-Diagnose)

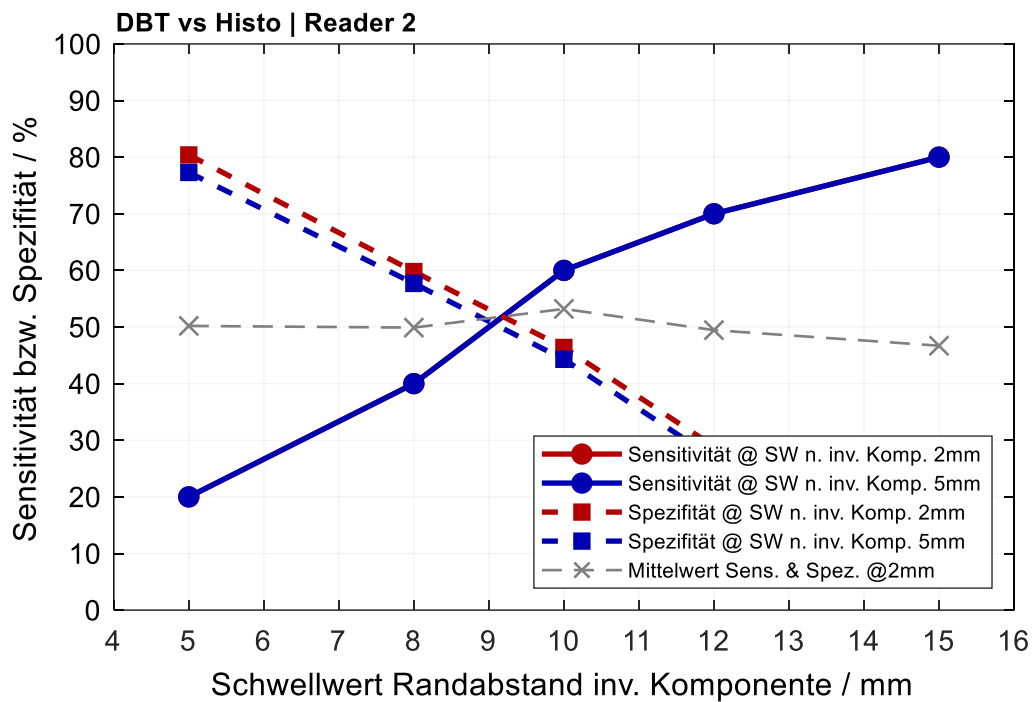


Abbildung 8.49: Sensitivitätsanalyse der R0/R1-Diagnose der DBT auf Randabstandskriterien (Reader 2/ Originale R0/R1-Diagnose)

8.7.6 Kontingenztafeln Erkennbarkeit Mikrokalk (Reader 2)

Tabelle 8.5: Kontingenztafel zum Vergleich der Erkennbarkeit von MK in der PR und der Histologie (Reader 2)

Kontingenztafel intraoperative PR vs Histologie Reader 2 Mikrokalk (abs.)

		PR Reader 2		Rand- häufigkeit		
		nein	ja			
Histo	nein	63	37	100	Spezifität	0,630
	ja	6	10	16	Sensitivität	0,625
Rand- häufigkeit		69	47	116		

Kontingenztafel intraoperative PR vs Histologie Reader 2 Mikrokalk (rel.)

		PR Reader 2		Rand- häufigkeit
		nein	ja	
Histo	nein	0,5431	0,31897	0,86207
	ja	0,051724	0,086207	0,13793
Rand- häufigkeit		0,59483	0,40517	

p0 0,62931

pe 0,56867

K2 0,14059

unterer Wert	-0,06318	oberer Wert	0,34437
---------------------	----------	--------------------	---------

Tabelle 8.6: Kontingenztabelle zum Vergleich der Erkennbarkeit von MK in der DBT und der Histologie (Reader 2)

Kontingenztabelle intraoperative DBT vs Histologie Reader 2 Mikrokalk (abs.)

		DBT Reader 2		Rand- häufigkeit		
		nein	ja			
Histo	nein	60	31	91	Spezifität	0,659
	ja	8	8	16	Sensitivität	0,500
Rand- häufigkeit		68	39	107		

Kontingenztabelle intraoperative DBT vs Histologie Reader 2 Mikrokalk (rel.)

		DBT Reader 2		Rand- häufigkeit
		nein	ja	
Histo	nein	0,56075	0,28972	0,85047
	ja	0,074766	0,074766	0,14953
Rand- häufigkeit		0,63551	0,36449	

p0 0,63551

pe 0,59499

K2 0,10006

unterer Wert -0,12509 **oberer Wert** 0,32522

8.8 Abhängigkeit der Güte der R0/R1-Diagnose von der Herdbefundgröße

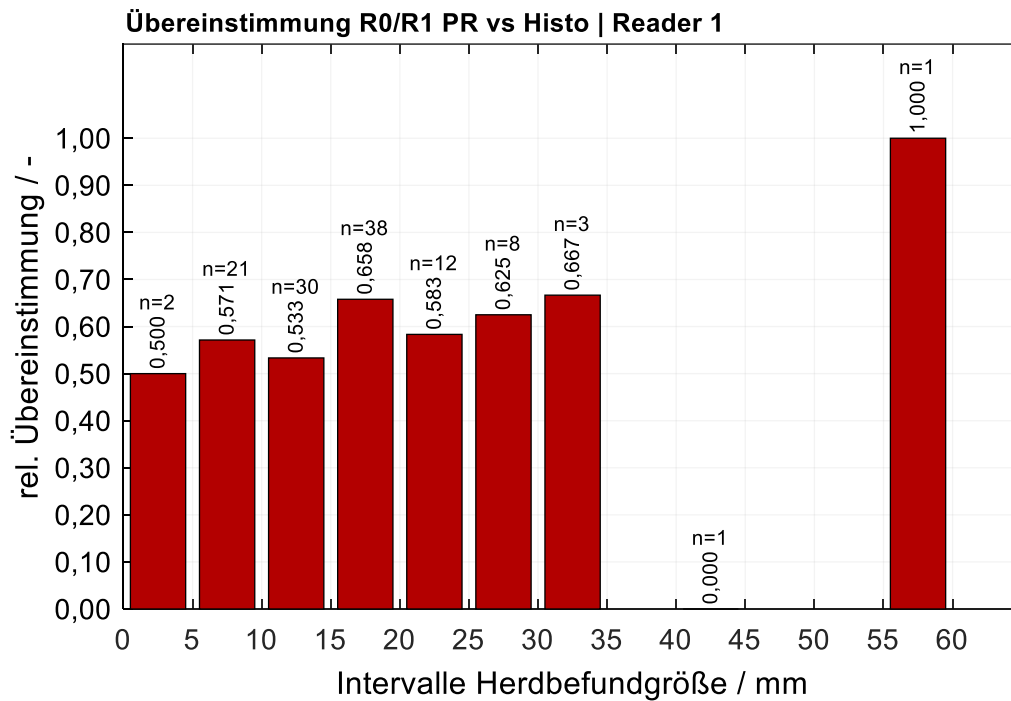


Abbildung 8.50: Relative Übereinstimmung der R0/R1-Diagnose aus der PR für Fälle innerhalb der jeweiligen Intervalle der Herdbefundgrößen (Reader 1)

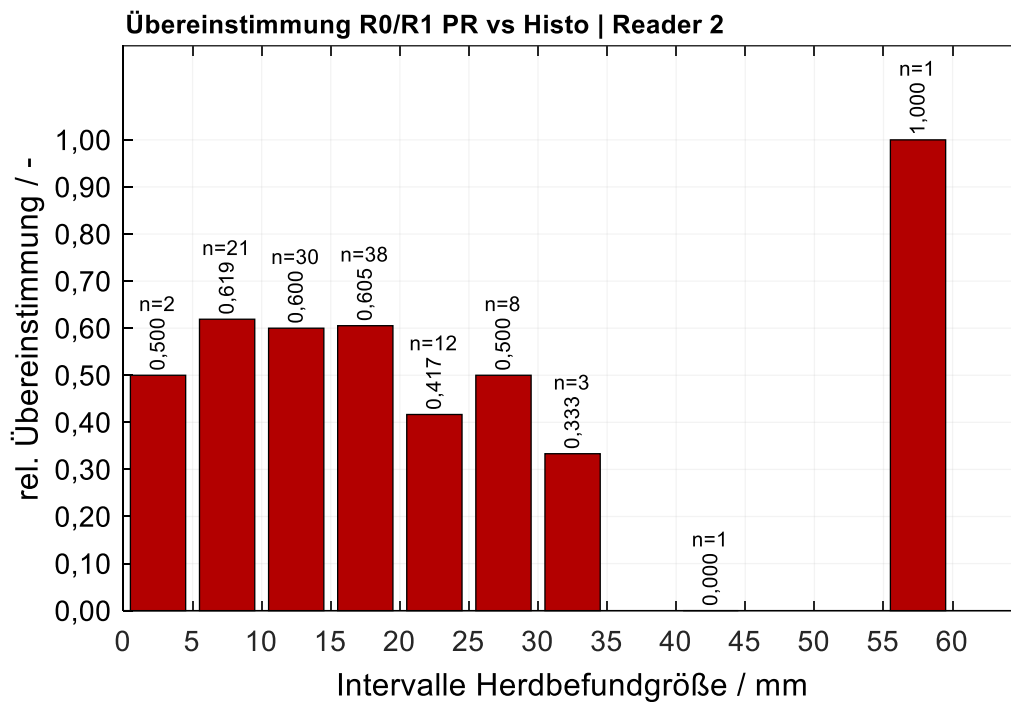


Abbildung 8.51: Relative Übereinstimmung der R0/R1-Diagnose aus der PR für Fälle innerhalb der jeweiligen Intervalle der Herdbefundgrößen (Reader 2)

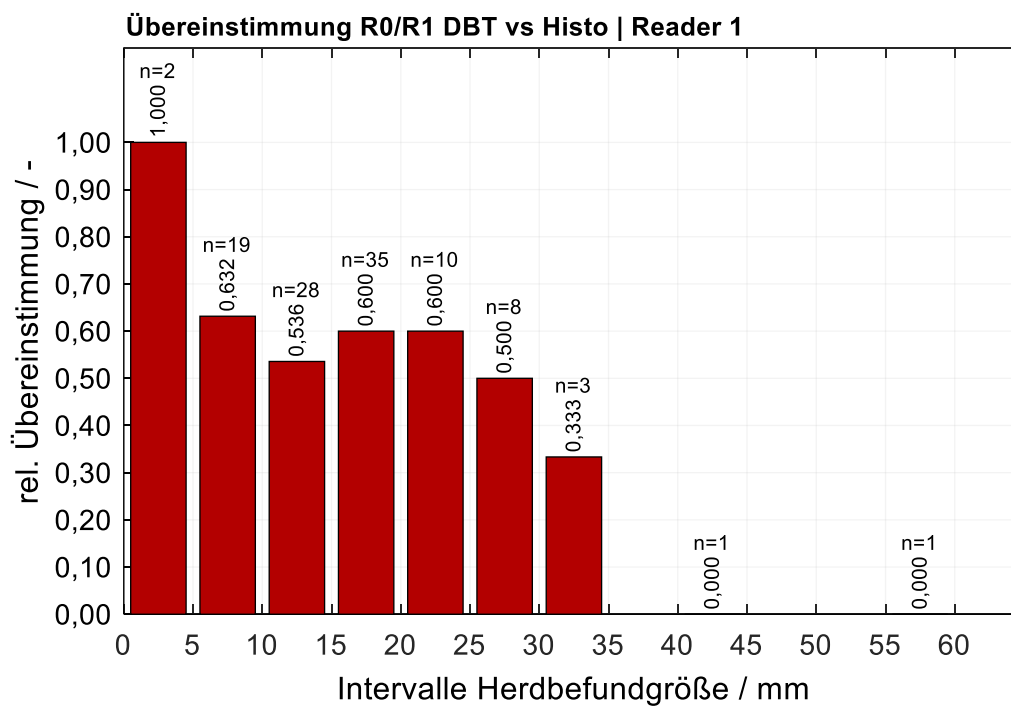


Abbildung 8.52: Relative Übereinstimmung der R0/R1-Diagnose aus der DBT für Fälle innerhalb der jeweiligen Intervalle der Herdbefundgrößen (Reader 1)

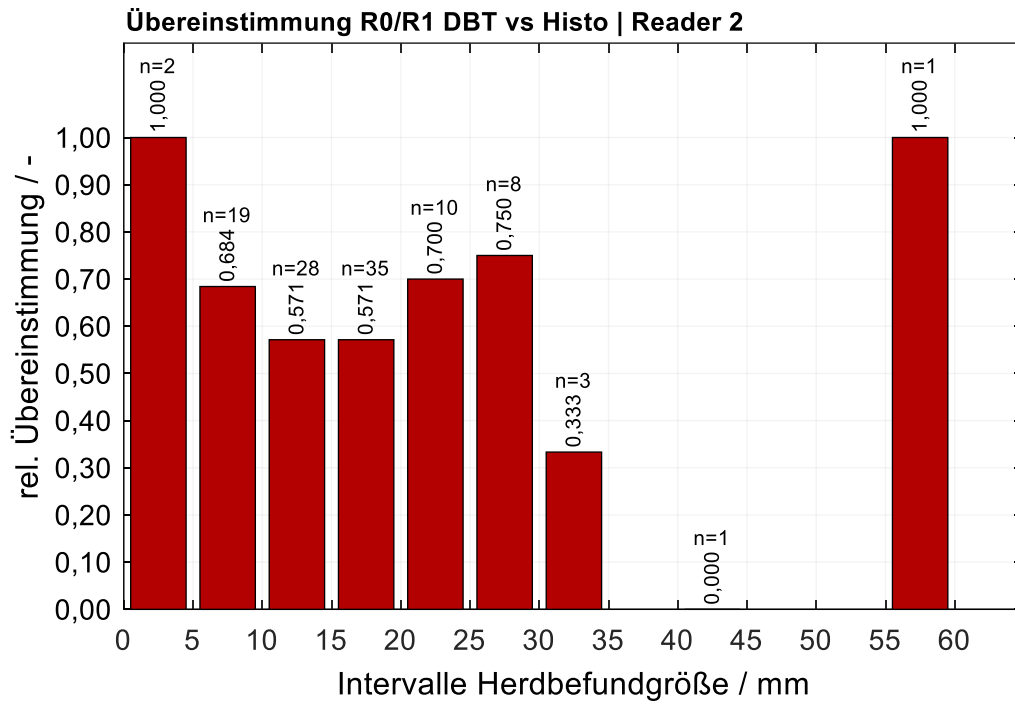


Abbildung 8.53: Relative Übereinstimmung der R0/R1-Diagnose aus der DBT für Fälle innerhalb der jeweiligen Intervalle der Herdbefundgrößen (Reader 2)