



## Ergebnisbericht (gemäß Nr. 14.1 ANBest-IF)

<b>Konsortialführung:</b>	Referenzzentrum Mammographie München
<b>Förderkennzeichen:</b>	01VSF18003
<b>Akronym:</b>	DIMASOS2
<b>Projekttitel:</b>	Dichte-indiziertes mammographisch-sonographisch Brustkrebs-Screening
<b>Autorinnen und Autoren:</b>	Sylvia H. Heywang-Köbrunner, Astrid Hacker, Eva Haußmann (Konsortialführung) Susann Elsner, Paula Grieger, Moritz Hadwiger, Alexander Katalinic (Konsortialpartner)
<b>Förderzeitraum:</b>	01. Juli 2019 - 30. Juni 2024
<b>Ansprechperson:</b>	Prof. Dr. med. Sylvia H. Heywang-Köbrunner Referenzzentrum Mammographie München, Sonnenstr. 29, 80331 München E-Mail: <a href="mailto:sylvia.heywang@referenzzentrum-muenchen.de">sylvia.heywang@referenzzentrum-muenchen.de</a> ; <a href="mailto:eva.haussmann@referenzzentrum-muenchen.de">eva.haussmann@referenzzentrum-muenchen.de</a> ; <a href="mailto:astrid.hacker@referenzzentrum-muenchen.de">astrid.hacker@referenzzentrum-muenchen.de</a> Tel: 089 961 661 386

Das dieser Veröffentlichung zugrundliegende Projekt DIMASOS2 wurde mit Mitteln des Innovationsausschusses beim Gemeinsamen Bundesausschuss unter dem Förderkennzeichen 01VSF18003 gefördert.

## Zusammenfassung

**Hintergrund:** Da dichtes Gewebe Tumore (CA) mammographisch verdecken kann, kann dieses CA-Entdeckungsrate und Effektivität des Mammographie (MX)-Screenings beeinträchtigen. Nutzen und Schaden eines kombinierten Mammographie-Sonographie-Screenings (MXUS) sowie dessen Integrierbarkeit ist für zentralisierte Programme noch unzureichend erforscht.

**Methodik:** In 15 Screeningeinheiten wurde die prospektiv, kontrollierte Mixed-Methods-Studie durchgeführt. Bei Frauen mit sehr hoher Drüsenkörperdichte (ca. obere 15 % DKD) wurde die Screening-MX durch eine ärztliche Ultraschall-Untersuchung (US) in der Interventionsgruppe (IG) ergänzt (MXUS-Screening). Die Kontrollgruppe (KG) bestand aus MX-Screening-Teilnehmerinnen (ScrTN) mit sehr hoher DKD derselben Screeningeinheiten. Ziel war es, die Anzahl und Stadienverteilung der durch MXUS zusätzlich entdeckten CA sowie die Rate zusätzlicher falsch-positiver (fpos) Empfehlungen zu bildgebender Abklärung (BA), Biopsie (BX) oder kurzfristigen Kontrolluntersuchungen (BI-RADS 3-Befunde) zu bestimmen. Es wurde für die gesundheitsökonomische Bewertung des ergänzenden US eine Budget-Impact Analyse mittels Markov-Modell modelliert. Umsetzbarkeit und Praktikabilität wurden durch qualitative Interviews mit ScrTN und Personal untersucht (Inhaltsanalyse nach Philipp Mayring).

**Ergebnisse:** In der Studie wurden in der IG 25.341 Frauen mit MXUS untersucht. Innerhalb dieser Gruppe wurde bei 2,7 % (95 %-KI: 2,1; 3,4) ScrTN zusätzliche CA im Vergleich zur alleinigen MX entdeckt. Im Vergleich zur unabhängigen KG ( $n = 38.539$ ) wurden in der IG insgesamt bei 3,45 % (KI: 1,90; 5,00) ScrTN zusätzliche CA gefunden. Die BA-Rate lag bei 6,8 % (versus 5,3 %), der die BX-Rate bei 3,2 % (versus 1,4 %). Die Tumorstadien waren in den beiden Gruppen ähnlich, die zusätzlich durch US gefunden CA waren etwas größer und seltener in situ CA. Die zusätzlichen Kosten pro ScrTN betrugen durchschnittlich ca. 45,59 €, vor allem durch den ergänzenden US mit Befundung, Folgeuntersuchungen und die potentiell zusätzlichen CA-Behandlungen. Würde bei ScrTN der oberen 15 % Dichte ein ergänzender US durchgeführt, entstünden den gesetzlichen Krankenkassen zusätzliche Kosten in Höhe von ca. 20,5 Millionen €. In den qualitativen Befragungen wussten etwa die Hälfte der ScrTN von ihrer hohen DKD und keine von der Beeinträchtigung der diagnostischen Genauigkeit, alle bewerteten die zusätzliche Untersuchung positiv und fühlten sich dadurch sicherer. Screening-ärzt\*innen überschätzten das Wissen der Frauen zur DKD, sahen die Notwendigkeit, die Sensitivität des Mammographie Screening Programms (MSP) für ScrTN mit hoher DKD zu erhöhen und nannten Hürden, wie unzureichende Evidenz bezüglich Sterblichkeitsreduktion, den Anstieg fpos Befunde sowie organisatorische Probleme.

**Diskussion:** Der verbesserten CA-Detektion durch MXUS steht eine erhöhte Rate fpos Befunde (Wiedereinbestellungsrate ca. +25 %) und die mehr als verdoppelte BX-Rate gegenüber bei durchschnittlich 45,59 € pro ScrTN mit hoher DKD. Gravierende organisatorische Hindernisse und fehlende Leitlinien erschweren die Implementierung der zusätzlichen Untersuchung.

**Schlagworte:** Mammographie; Brustultraschall; Früherkennung von Krebs; Risikoabschätzung; Programmevaluierung; Brustdichte; Mammatumoren; Budget-Impact Analyse; Qualitative Befragung; automatisierte Brustdichtemessung

## Inhaltsverzeichnis

I	Abkürzungsverzeichnis .....	6
II	Abbildungsverzeichnis .....	7
III	Tabellenverzeichnis .....	7
1	Projektziele .....	8
1.1	Hintergrund .....	8
1.2	Ziele der Studie .....	9
2	Projektdurchführung .....	10
2.1	Projektbeteiligte .....	10
2.2	Beschreibung/ Darstellung des Projekts .....	12
2.3	Beschreibung Ablauf des Projekts .....	14
2.4	Erfahrungen mit der Implementierung/ Maßnahmen .....	15
2.5	Dichtebestimmung .....	16
2.6	Rechtsgrundlage .....	17
3	Methodik .....	18
3.1	Quantitativer Studienteil .....	18
3.1.1	Studiendesign .....	18
3.1.2	Studienpopulation .....	18
3.1.2.1	Bildung der Studiengruppe .....	18
3.1.2.2	Bildung der Kontrollgruppe .....	19
3.1.3	Fallzahlplanung .....	19
3.1.4	Untersuchungsmethode .....	20
3.1.4.1	Das MXUS-Verfahren .....	20
3.1.4.2	Medizinischer Ablauf bis Konsensusbefundung/ DIMASOS2-Screeningkette .....	21
3.1.4.3	Befundungsergebnisse in der Interventionsgruppe .....	21
3.1.4.4	Nomenklatur .....	22
3.1.4.5	Medizinischer Ablauf ab bildgebender Abklärung .....	22
3.1.4.6	Dokumentation der Daten .....	23
3.1.5	Automatisierte Dichtebestimmung .....	24
3.1.6	Qualitätssicherung und Monitoring .....	24
3.1.7	Statistische Auswertung .....	25
3.1.7.1	Datenaufbereitung bei der Studienleitung .....	25
3.1.7.2	Fragestellung .....	26
3.1.7.3	Auswertung der Daten/ Statistische Methoden .....	26
3.2	Gesundheitsökonomische Analyse .....	28

3.3	Qualitativer Studienteil .....	30
4	Projektergebnisse .....	32
4.1	Ergebnisse des quantitativen Studienteils .....	32
4.1.1	Voraussetzung Dichtemessung.....	33
4.1.2	Deskriptive Beschreibung der Studienpopulation.....	35
4.1.3	Karzinom-Entdeckungsrate für MX und MXUS, gesamt und stratifiziert nach Subgruppen (Dichte und Alter) .....	35
4.1.4	Beschreibung und Vergleich der entdeckten CA .....	35
4.1.5	Zusätzliche bildgebende und histologische Abklärungen.....	36
4.1.6	Abschätzung der Sensitivität und Spezifität der Studienmethode, dem kombinierten Verfahren aus MX und US (MXUS) .....	37
4.1.7	Nutzen/Schaden Verhältnis in Abhängigkeit von der Dichte .....	38
4.1.8	Abhängigkeit der CA-entdeckungsrate durch MXUS je nach gesetztem Schwellenwert	39
4.2	Ergebnisse des gesundheitsökonomischen Studienteils.....	39
4.3	Ergebnisse des qualitativen Studienteils.....	40
4.3.1	Wahrnehmung und Kenntnisse zum Thema sehr hohe DKD .....	40
4.3.2	Einfluss DKD auf die diagnostische Genauigkeit des MSP .....	41
4.3.3	Durchführbarkeit und Praktikabilität von MXUS .....	41
5	Diskussion der Projektergebnisse.....	42
5.1	Quantitativer Studienanteil.....	42
5.1.1	Kohorte und Studienpopulation .....	42
5.1.2	Auswahl des zu untersuchenden Dichtebereichs (sehr hohe DKD, obere 15 %) .....	42
5.1.3	Wahl der Studienmethode .....	44
5.1.4	Primäre Fragestellung: Karzinom-Mehrentdeckung .....	44
5.1.5	Beschreibung der zusätzlich gefundenen CA (Klassifikation, CA-Stadienverteilung) .....	46
5.1.6	Beschreibung der falsch positiven Befunde (Wiedereinbestellungsrate, Biopsierate, Rate an kurzfristigen Kontroll-Empfehlungen (sog. BI-RADS 3-Befunde) (Erwartet war eine Biopsierate < 5 %.) .....	47
5.1.7	Ergänzende Untersuchungen zu Nutzen und Risiko, stratifiziert nach dem Dichtebereich (obere 5/ 10/ 15 %) MXUS.....	49
5.1.8	Klinischer Nutzen der Studienmethode.....	50
5.2	Gesundheitsökonomisches Modell.....	51
5.3	Qualitativer Studienteil .....	51
5.3.1	Wahrnehmung und Kenntnisse zum Thema DKD .....	52
5.3.2	Einfluss von hoher DKD auf die diagnostische Genauigkeit im MSP .....	52
5.3.3	Durchführbarkeit und Praktikabilität von MXUS .....	53
5.3.4	Stärken und Schwächen der qualitativen Studie.....	53

5.4	Zusammenfassende Bewertung der Ergebnisse .....	53
5.4.1	Dichthemessung und -stratifizierung.....	53
5.4.2	Zusammenfassende Wertung der medizinischen Ergebnisse .....	55
5.4.3	Machbarkeit.....	58
5.4.4	Stärken und Schwächen der Gesamt-Studie: .....	59
6	Verwendung der Ergebnisse nach Ende der Förderung.....	61
7	Erfolgte bzw. geplante Veröffentlichungen .....	62
7.1	Geplante Veröffentlichungen.....	62
7.2	Folgende Veröffentlichungen sind noch geplant, bzw. in Vorbereitung: .....	62
IV	Literaturverzeichnis .....	63
V	Anlagen .....	67

## I Abkürzungsverzeichnis

ACR	American College of Radiology
AG	Altersgruppe
BA	Bildgebende Abklärung
BIA	Budget-Impact-Analyse
BI-RADS	Breast Imaging Reporting and Data System
BMV-Ä	Bundesmantelvertrag- Ärzte
BX	Biopsie
CA	Karzinom, Tumor
DCIS	Duktal Carcinoma in situ
DIMASOS2	Dichte-indiziertes mammographisch-sonographisches Brustkrebs-Screening
DKD	Drüsenkörperfiedliche
EB	Erstbefund/ Ersbefunder*in
fneg	falsch-negativ
fpos	falsch-positiv
GKV	gesetzliche Krankenversicherung
IA	Innovationsausschuss
IG	Interventionsgruppe
ISE	Institut für Sozialmedizin und Epidemiologie
IPTW	Inverse Propensity Treatment Weighting
KG	KG
KI	Künstliche Intelligenz
KK	Konsensuskonferenz/ Konsensus-Befundung
MISP	Mammographie-Screening-Programm
M-Status	Metastasen-Status
MIX	Mammographie
MXUS	Mammographie plus Ultraschall
N-Status	Lymphknotenstatus
PACS	Picture Archiving and Communication System (Patient*innen Archivierungs-System)
PVÄ	Programverantwortliche-Ärztin/ Programverantwortliche-Arzt/ Programverantwortliche-Ärzt*innen
RFA	Radiologische Fachkraft
ScrTN	Screening-Teilnehmerin/ -Teilnehmerinnen (am MSP)
SE	Screening-Einheit(en)
TN	Teilnehmerin/ Teilnehmerinnen
T-Status	Tumor-Status
US	Ultraschall (handgehalten)
ZB	Zweitbefund/ Zweitbefunder*in

## **II      Abbildungsverzeichnis**

Abbildung 1: Markov-Modell des Budget-Impact Modells mit Übergangswahrscheinlichkeiten p1 und p2 („US“: ergänzende US-Untersuchung mit MXUS-Befundung = Studienmethode) .....	29
Abbildung 2: Induktive Kodierung der Interviewdaten in der DIMASOS2-Studie (Ärzt*innen & Screeningteilnehmerinnen) .....	32

## **III      Tabellenverzeichnis**

Tabelle 1: Projektbeteiligte Personen .....	10
Tabelle 2: Beschreibung der Endpunkte (Outcomes) und die Art der jeweiligen Erfassung... 27	

## 1 Projektziele

### 1.1 Hintergrund

Das offizielle Mammographie-Screening Programm (MSP) in Deutschland zur Früherkennung von Brustkrebs richtet sich an Frauen im Alter zwischen 50 und 75 Jahren. Während der Studiendauer betrug die Alterspanne 50 bis 69 Jahre; die Anhebung des Screeningalters erfolgte im Juli 2024 nach Beendigung dieser Studie.

Die Effektivität von Brustkrebs Screening Programmen wird bezogen auf die eingeladene Frau, mit einer relativen Mortalitätsreduktion für Brustkrebs von rund 20 % (für die tatsächlich teilnehmende Frau liegt der Effekt höher) angegeben, welche auf die Entdeckung von Brustkrebs in früheren Stadien, einem damit einhergehenden früheren Therapiebeginn und einer besseren Prognose zurückzuführen ist (Broeders 2012; Marmot 2013). Für das deutsche MSP berichteten Katalinic et al. (2020) eine Reduktion der fortgeschrittenen Stadien und Waldmann et al. (2021) eine Mortalitätsreduktion von 23,6 % für Frauen im Alter zwischen 60 und 69 Jahren und von 28,8 % für Frauen im Alter zwischen 50 und 59 Jahren.

Für Frauen mit hoher Drüsengewebsdichte (DKD) ist eine geringere Effektivität zu erwarten, denn dichtes Gewebe kann maligne Veränderungen der Brust überdecken, sodass Brustkrebs bei der Befundung manchmal unentdeckt bleibt. Eine mit Mammographie kombinierte Ultraschalluntersuchung (MXUS) kann auch bei asymptomatischen Frauen mit hoher DKD die Sensitivität im Vergleich zur alleinigen Mammographie (MX) steigern (Rebolj 2018; Corsetti 2011; Wu 2022). In einer kürzlich erschienenen Metaanalyse betrug die für Ultraschall (US) berichtete Mehrentdeckungsrate bei asymptomatischen Frauen mit hoher DKD ca. 3 pro 1.000 (Glechner 2023).

In Deutschland, wie in bislang auch international in zentralisierten Screening-Programmen, gibt es für Frauen mit hoher DKD keine von der Normalpopulation abweichende Screening-Empfehlung. Zusätzliche Untersuchungen, wie z. B. US, werden zur weiteren Beurteilung von mammographisch oder klinisch festgestellten, unklaren Befunden eingesetzt (Dench 2020; Albert 2008). In der Regel werden Frauen, die am MSP teilnehmen, nicht über ihre DKD informiert oder erfahren bei hoher DKD nicht, dass die diagnostische Genauigkeit der MX beeinträchtigt sein kann.

Potenzielle Hindernisse für eine evtl. Einführung von zusätzlichem US in zentralisierten Screeningprogrammen sind u. a. der Mangel an Ressourcen für Durchführung und Befundung einer qualitätsgesicherten MXUS-Untersuchung wegen der vielfach längeren Untersuchungszeit für US im Vergleich zur MX und wegen der schwierigen Qualitätssicherung des US bei hoher Untersucher-Abhängigkeit (Wöckel 2018). Hinzu kommen eine deutlich höhere Zahl an Biopsie-Empfehlungen durch US, die damit sowie mit ergänzender Bildgebung und Wiedereinbestellung verbundene psychische Belastung der betroffenen Frauen und die erhöhte Arbeitsbelastung der Leistungserbringenden. Der Kosten-Nutzen-Effekt der zusätzlichen Leistung ist unklar (Lauby-Secretan 2015). Aus diesen Gründen hängt die Durchführbarkeit einer zusätzlichen Untersuchung von einer sorgfältigen Auswahl der Frauen mit dem höchsten Risiko unter den Frauen mit hoher DKD ab (Lobig 2023).

Im Rahmen von DIMASOS2 wurde in 15 Screeningeinheiten (SE) in Deutschland ein zusätzlicher US für Frauen mit hoher DKD angeboten, um die Effizienz und die Praktikabilität des Verfahrens zu evaluieren. Es handelte sich dabei um eine Mixed-Methods-Studie mit einem quantitativen und einem qualitativen Studienteil sowie einer Budget-Impact-Analyse (BIA) zur Modellierung der zusätzlichen Kosten aus Sicht der gesetzlichen Krankenversicherung (GKV). Als Vergleichsgruppe dienten Frauen mit hoher DKD, denen kein zusätzlicher US angeboten wurde (herkömmliches MSP).

## 1.2 Ziele der Studie

- (1) Bestimmung des Effekts des MXUS auf eine zusätzliche Karzinom -Entdeckung (primäre Fragestellung, Erwartung: 2 / 1000 zusätzlich entdeckte Karzinome (CA)).
- (2) Bestimmung der Sensitivität einer bei sehr hoher DKD im MSP eingesetzten kombinierten Untersuchung (MXUS) im Vergleich zur alleinigen MX.
- (3) Beschreibung der zusätzlich gefundenen CA (Klassifikation, CA-Stadienverteilung).
- (4) Beschreibung der falsch-positiv (fpos) Befunde (Wiedereinbestellungsrate, BX-Rate, Rate der kurzfristigen Kontrollempfehlungen; wobei letztere BI-RADS 3-Befunden nach bildgebender Abklärung (BA) entsprechen. Erwartet war eine BX-Rate < 5 %).
- (5) Ergänzend sollte untersucht werden, in welchem der sehr hohen Dichte-Bereiche (obere 5/ 10/ 15 % der gemessenen Dichtewerte) MXUS unter Berücksichtigung von Nutzen und Risiko optimal einsetzbar ist.
- (6) Gesundheitsökonomische Analyse zur Bestimmung des Kosten-Nutzen Verhältnisses. (durch den Konsortialpartner)
- (7) Befragung von ärztlichem Personal und RFA zur Machbarkeit und Relevanz des zusätzlichen US und Befragung von Teilnehmerinnen (TN) der IG und des Screening-Personal zu Relevanz und Wissen zur DKD sowie Akzeptanz von MXUS. (qualitativer Studienteil des Konsortialpartner)

## 2 Projektdurchführung

### 2.1 Projektbeteiligte

Tabelle 1: Projektbeteiligte Personen

<b>1. Projektleitung</b>		
Prof. Dr. med. Sylvia H. Heywang-Köbrunner	Konsortialführung	Referenzzentrum Mammographie, Sonnenstraße 29, 80331 München, Tel.: 089 961 661 330, E-Mail: <a href="mailto:sylvia.heywang@referenzzentrum-muenchen.de">sylvia.heywang@referenzzentrum-muenchen.de</a>
Dr. med. Astrid Hacker	Wiss. Mitarbeit	Referenzzentrum Mammographie, Sonnenstraße 29, 80331 München, Tel.: 089 961 661 342, E-Mail: <a href="mailto:astrid.hacker@referenzzentrum-muenchen.de">astrid.hacker@referenzzentrum-muenchen.de</a>
Eva Haußmann	Projektkoordination	Referenzzentrum Mammographie, Sonnenstraße 29, 80331 München, Tel: 089 961 661 341, E-Mail: <a href="mailto:eva.haussmann@referenzzentrum-muenchen.de">eva.haussmann@referenzzentrum-muenchen.de</a>
Emanuela Trifan	Dokumentation	Nicht mehr am RZM beschäftigt
Souhir Abid	Dokumentation	Referenzzentrum Mammographie, Sonnenstraße 29, 80331 München, Tel: 089 9961 661 344, E-Mail: <a href="mailto:souhir.abid@referenzzentrum-muenchen.de">souhir.abid@referenzzentrum-muenchen.de</a>
Andreas Baar	Projektmanagement	Nicht mehr am RZM beschäftigt
Andrea Meier	Projektmanagement	Referenzzentrum Mammographie, Sonnenstraße 29, 80331 München, Tel: 089 9961 661 331, E-Mail: <a href="mailto:andrea.meier@referenzzentrum-muenchen.de">andrea.meier@referenzzentrum-muenchen.de</a> <a href="https://@referenzzentrum-muenchen.de">https://@referenzzentrum-muenchen.de</a>
<b>2. Konsortialpartner</b>		
Prof. Dr. med. Alexander Katalinic	Evaluation	Institut für Sozialmedizin und Epidemiologie, Ratzeburger Allee 160, 23562 Lübeck, Tel.: 0451 500 51201, E-Mail: <a href="mailto:Alexander.katalinic@uksh.de">Alexander.katalinic@uksh.de</a>

Dr. Susanne Elsner	Wiss. Mitarbeit	Institut für Sozialmedizin und Epidemiologie, Ratzeburger Allee 160, 23562 Lübeck, Tel: 0451 500 51213, E-Mail: susanne.elsner@uksh.de
Paula Grieger	Wiss. Mitarbeit	Institut für Sozialmedizin und Epidemiologie, Ratzeburger Alle 160, 23562 Lübeck, Tel: 0451 500 51229, E-Mail: paula.grieger@uksh.de
Dr. Moritz Hadwiger	Wiss. Mitarbeit	Nicht mehr am ISE beschäftigt
Dr. med. Christine Matthis	Wiss. Mitarbeit	Institut für Sozialmedizin und Epidemiologie, Ratzeburger Alle 160, 23562 Lübeck, Tel. 0451 500 51201, E-Mail: christine.matthis@uksh.de
Andrea Rieck	Dokumentation	Institut für Sozialmedizin und Epidemiologie, Ratzeburger Allee 160, 23562 Lübeck, Tel.: 0451 500 51283, E-Mail: andere.rieck@uksh.de
Katharina Block	Dokumentation	Nicht mehr am ISE beschäftigt

Die Leitung des Projektes sowie die Konsortialführung lagen bei Prof. Dr. med. Sylvia Heywang-Köbrunner, Leiterin des Referenzzentrums Mammographie München. Weitere projektbeteiligte Personen in München waren: Dr. med. Astrid Hacker (wissenschaftliche Mitarbeit, Unterstützung Projektleitung und Projektkoordination), Eva Haußmann (Projektkoordination und wissenschaftliche Mitarbeit), Emanuela Trifan (Dokumentation), Souhir Abid (Dokumentation), Andreas Baar (Projektmanagement und Verwaltung) und Andrea Meier (Projektmanagement und Verwaltung).

Konsortialpartner war Professor Dr. med. Alexander Katalinic, Institut für Sozialmedizin und Epidemiologie (ISE), Universität Lübeck. Weitere projektbeteiligte Personen aus dem ISE waren: Dr. Susanne Elsner (wissenschaftliche Mitarbeit, Projektkoordination, Evaluation, qualitativer Studienteil), Dr. Moritz Hadwiger (Evaluation), Paula Grieger (Evaluation, qualitativer Studienteil), Dr. Christine Matthis (Evaluation), Andrea Rieck (Dokumentation, qualitativer Studienteil), Katharina Block (Dokumentation).

MammaSoft und KV-IT programmierten eine Anpassung der nationalen Screening-Softwares (MammaSoft und MaSc) für den Datenaustausch zwischen Studienserver und Screening-Software. Die kanadische Firma Densitas stellte die Dichtesoftware für die automatische Dichtebestimmung zur Verfügung. Die Fa. IdeeDesign (Michael Hertlein) programmierte die Studiensoftware und sorgte für die Anbindung der peripheren Studienserver der SE an den zentralen Studienserver bei der Projektleitung, sowie für den datenschutzgerechten Datenaustausch zwischen Studienserver und nationalen Screening-Softwares MammaSoft und MaSc. Sie unterstützte zudem alle Datenexporte sowie die Vorbereitung zur statistischen

Evaluation. Die Studienzentren rekrutierten bundesweit TN für die Interventionsgruppe (IG) und Kontrollgruppe (KG).

## 2.2 Beschreibung/ Darstellung des Projekts

Bei der Zielgruppe handelte es sich um TN des MSP mit sehr hoher DKD (definitionsgemäß mindestens die oberen 15 % der in der Population gemessenen Dichtewerte). Die DKD wurde mit einer in den Untersuchungsablauf integrierten Software densitasdensity™ der Firma Densitas (Halifax, Kanada) für jede MX der teilnehmenden Screeningstandorte automatisiert gemessen. Geeigneten Frauen mit sehr hoher DKD, deren MX zu vorgegebenen Studienzeiten durchgeführt wurde, wurde direkt nach Erstellung der Screening-MX angeboten, zusätzlich einen US durchzuführen zu lassen. Der Termin für den Ultraschall wurde am selben Tag oder spätestens innerhalb von 7 weiteren Kalendertagen nach der MX angesetzt. Die Auswertung des US erfolgte immer durch Kombination der Informationen aus Mammographie und Sonographie (MXUS).

Frauen, die entsprechend ihrem zufällig in die Studienzeit fallenden MX-Screening-Termin einem ergänzenden US (mit MXUS-Befundung) zustimmten und diese Untersuchung erhielten (IG), wurden zudem mit einer Kontrollgruppe (KG) aus Screeningteilnehmerinnen (ScrTN) verglichen, bei denen ebenfalls eine sehr hohe DKD festgestellt wurde, die jedoch kein Angebot für einen zusätzlichen US erhalten haben (KG).

Die primäre Hypothese der Studie lautete, dass durch die Kombination von zusätzlichem US und MX bei 1000 gescreenten Frauen mit sehr hoher DKD mindestens zwei zusätzliche CA entdeckt werden können.

Einschlusskriterien waren:

- TN des MSP
- Alter zwischen 50 und 69 Jahre
- Sehr hohe DKD (abhängig vom Densitas-Score)
- Informierte schriftliche Einwilligung

Ausschlusskriterien waren:

- Fehlende Einwilligung (oder Einwilligungsfähigkeit)
- große Fremdkörper und Brustimplantate
- Teilnahme an einer anderen diagnostischen Studie

Der zusätzliche Zeitaufwand für den US mit Dokumentation nach kombinierter Bewertung von US und MX inklusive MX-Voraufnahmen betrug ca. 20-30 Minuten. Vor Studienbeginn war für die beteiligten Screening-Ärzt\*innen aus den 15 teilnehmenden SE ein 2-tägiger Workshop zum Thema "Qualitätsgesicherte MXUS-Untersuchung" obligat. Dieser wurde von der Studienleitung mit Unterstützung der DEGUM durchgeführt.

Die Qualitätssicherung umfasste auch die US-Geräte, die Eignung der voll ausgebildeten Screening-Befunder\*innen anhand vorgegebener Vorerfahrungen mit Brust-US, die Durchführung der kombinierten MXUS-Befundung (statt alleiniger US-Befundung), 2-tägige

Schulung zu Technik, Befundung, Standards und Pitfalls sowie kontinuierliches Feedback zu eigenen medizinischen Leistungsparametern (Benchmark).

Wie im MSP wurden eine unabhängige Doppelbefundung, sowie bei allen Auffälligkeiten zusätzlich die Konsensusbefundung (KK) sowohl für MX als auch für MXUS prospektiv systematisch umgesetzt. Dieses Prozedere, inklusive der Verblindung der Erstbefunde (EB) für die Zweitbefunder (ZB), wurde auch durch die Studiensoftware garantiert.

Die abschließenden Befundungs-Ergebnisse (nach MXUS) wurden von der Studien-Software prospektiv an die Screenings-Softwares (MammaSoft und MaSc) übertragen. Der Ablauf der Screening-Kette blieb also unberührt und identisch zum MSP. Identisch zum MSP erfolgten bei allen histologisch gesicherten Befunden die Standard-Screening-Konferenzen (prä- und post-operative Multidisziplinäre Fallkonferenzen) sowie die Ergebniserfassungen in den Screeningsoftwares.

Die Vergütung des zusätzlichen US, der zusätzlichen Befundungen und zusätzlichen Folgeuntersuchungen erfolgten in der Studie angelehnt an den EBM, ebenfalls unterstützt durch die Studien-Software.

Im Rahmen der gesundheitsökonomischen Analyse des Konsortialpartner wurde ein Markov-Modell erstellt, in dem ein Entscheidungsbaum verschachtelt ist, um die ökonomischen Auswirkungen der Intervention zu modellieren. Der Entscheidungsbaum wurde dazu verwendet, die herkömmlichen Kosten des MSP mit den zusätzlichen Kosten für den US und damit verbundene potentielle Folgeuntersuchungen (z. B. weitere bildgebende Verfahren = bildgebende Abklärung (BA) oder Biopsien (BX)) sowie den zu erwartenden gesundheitlichen Nutzen in Relation zu setzen. Zum erwarteten Nutzen der kombinierten Untersuchung zählen: frühere Diagnosen und niedrigere Tumorstadien sowie damit verbundene bessere Prognosen, die sich langfristig als geringere ökonomische Belastung auswirken können.

Einige Frauen (n = 22) aus der IG nahmen zudem an einem semi-strukturierten Telefoninterview teil. Die TN wurden vom Screeningpersonal aus sechs SE rekrutiert. Die Interviews wurden von Mitarbeiter\*innen des ISE vorbereitet und durchgeführt. Im Rahmen eines Telefoninterviews wurden TN darüber befragt, ob sie vor der Studienteilnahme bereits wussten, dass sie eine hohe DKD haben, welche Erfahrungen sie mit der zusätzlichen Untersuchung gemacht haben und ob, und wenn ja, sich ihre Einstellung gegenüber dem MSP verändert hat. Die Interviews dauerten im Schnitt 20 Minuten und TN erhielten 50 Euro Aufwandsentschädigung.

Außerdem wurden Screening-Ärztinnen (n = 7) und RFA (n = 2) zu ihren Erfahrungen mit der zusätzlichen Untersuchung befragt. Die Interviews fanden zwischen 2020 und 2021 statt, folgten einem semi-strukturierten Interviewleitfaden und wurden telefonisch von Mitarbeiterinnen des ISE durchgeführt. Die Teilnahme an einem Interview wurde mit einer Aufwandsentschädigung von 50 Euro vergütet.

Die beiden Interviewleitfäden für die Durchführung der Telefoninterviews (Studie des Konsortialpartner) einer für ScrTN und einer für das Screeningpersonal – wurden auf der Grundlage einer umfassenden Literaturrecherche entwickelt. Sie wurden im gesamten Team in mehreren Treffen diskutiert, angepasst und schließlich in einem Vortest auf ihre praktische Anwendbarkeit hin getestet.

## 2.3 Beschreibung Ablauf des Projekts

Die Studie wurde von der Ethikkommission der Universität zu Lübeck beraten und befürwortet, begann am 01.07.2019 und endete nach Verlängerung am 30.06.2024.

Nach Vertragsabschluss mit den Programmverantwortlichen-Ärzt\*innen (PVÄ) der teilnehmenden SE erfolgte die Installation der Dichtesoftware sowie der Studiensoftware, die die offiziellen Screeningprogramme (MammaSoft und MaSc) mit der Datenbank zur Erfassung der DIMASOS2-Ergebnisse verband. Erst nach abgeschlossener Vorbereitungsphase und Schulung der Teams konnte pro SE in ursprünglich 21 Zentren in 19 SE mit der Rekrutierung begonnen werden. Der Rekrutierungsstart an den unterschiedlichen Standorten erstreckte sich auf einen Zeitraum von Mai 2020 bis Juni 2021. Leider starteten 4 SE (5 Zentren) nach abgeschlossener Vorbereitungsphase nicht mit der TN-Rekrutierung. Begründet wurde dieses jeweils durch Probleme durch die Corona-Epidemie, Personalmangel und durch unzureichende Finanzierung der zu erbringenden Leistungen bezogen auf deren Aufwand.

Zu Beginn der Studie war geplant, die IG mit einer historischen Kohorte zu vergleichen. Hierfür wäre es erforderlich gewesen, die DKD retrospektiv an ca. 200.000 digitalen MX der teilnehmenden Standorte zu bestimmen, um eine KG von 30.000 Frauen mit hoher DKD (obere 15 %) zu ermitteln. Hier ergaben sich unerwartete Probleme auf Grund der in den SE sehr unterschiedlichen Bild-Archivierungssysteme (PACS) und der Nicht-Zulässigkeit eines unterstützenden Zugriffs der Studienleitung auf fremde PACS-systeme. Hinzu kam, dass ein Teil dieser MX an inzwischen ersetzen älteren (z. T. auch veralteten) MX-Geräten erstellt waren. Eine Adaptierung des Dichte-Programms auf diese MX-Technik war nicht mehr sinnvoll.

Parallel zeigte sich, dass keines der teilnehmenden Zentren die erwartete Rekrutierungszeit auf Grund von personellen Engpässen und der Corona-Pandemie zur Verfügung stellen konnte. Hier bot sich an, die KG aus Untersuchungen zu bilden, für die Dichtemessungen außerhalb der Rekrutierungszeit an den teilnehmenden Standorten vorlagen, also in Zeiten von „kein Angebot“. Diese KG hatte im Vergleich zur historischen Kohorte folgende Vorteile: dieselbe Gerätetechnik, dieselben Ersteller\*innen- und Befunder\*innen-Teams, vergleichbare Populationen, vergleichbarer Untersuchungszeitraum. Bei datenschutzrechtlich unproblematischer Dichteermittlung und Anonymisierung von Einzelfiles wurde das Amendment auch von der Ethikkommission unterstützt.

Sowohl die Installation der Studien-Hard- und Software, die Implementierung der Studien-Abläufe als auch der Beginn der Rekrutierung verzögerten sich um mehrere Monate bedingt den Corona-Shutdown und hieraus resultierende Ausfälle beim technischen, medizinischen und organisatorischen Personal der teilnehmenden SE.

Auch bei der Rekrutierung ergaben sich weitere erhebliche Verzögerungen. Diese waren bedingt einerseits durch die Corona-Pandemie mit Shutdown. Ein Rekrutierungsstop über > 6 Wochen führte zu einem Nachholbedarf an Screening-Untersuchungen. So mussten die SE unter schwierigen Umständen (wechselnde Hygiene-Vorschriften, erhöhter psychischer und körperlicher Stress beim Personal) vermehrt screeen und andererseits zunehmende Engpässe an Fachpersonal kompensieren (Personalausfälle durch Krankheit, Kinderbetreuung

und Quarantäne-Vorschriften sowie Personalmangel, u. a. durch Abwanderung von medizinischem Personal).

Erwartbare Verzögerungen ergaben sich durch die geringere Anzahl an teilnehmenden Zentren. Diese Zahl konnte trotz intensivster Bemühungen und persönlicher Kontaktaufnahme zu fast allen PVÄ der 96 bundesweiten SE nicht erhöht werden.

Auch die Rekrutierung in den SE erwies sich als deutlich schwieriger als erwartet, ebenfalls auf Grund personeller Engpässe, der langen Dauer der US-Untersuchungen bei der geforderten hohen Expertise der MXUS-Befunder und durch Motivationseinbußen bei der (eigentlich erwartbar) niedrigen Anzahl an zusätzlich auffindbaren richtig positiven Befunden bei der ergänzender US asymptomatischer Frauen (Screening-Situation). Von mehreren Verwaltungen teilnehmender oder potentiell teilnehmender SE wurde die zusätzliche US-Untersuchung bei der gegebenen Studienerstattung als nicht-kostendeckend angesehen und damit nicht unterstützt.

Die oben genannten Gründe führten zu 3 Laufzeitverlängerungen der Studie, sowie zur Fallzahlreduktion von 30.000 auf 25.000 TN in der IG.

Trotz der Reduzierung der Fallzahl, die die Power der Studie beeinträchtigt, konnten weiterhin zuverlässige Effektschätzer ermittelt werden, wobei jedoch eine Präzisionsreduktion hierdurch von etwa 11 % kalkuliert wurde. Die finalen Gruppen umfassen 25.341 Frauen in der IG und 38.529 in der KG.

## 2.4 Erfahrungen mit der Implementierung/ Maßnahmen

Die Rekrutierung der teilnehmenden SE sowie die Installation der Dichte-Software wurden durch die Konsortialführung organisiert. Eine ursprünglich geplante Anzahl von mehr als 20 teilnehmenden SE konnte nicht erreicht werden. Hauptgründe für eine Nichtteilnahme von SE, deren Ärzte Interesse bekundet hatten, waren einerseits die Einschätzung in verschiedenen SEs, dass eine Studienteilnahme nicht kostendeckend erfolgen würde, andererseits die mangelnden personellen Ressourcen für den zeitintensiven US in den SE. Am Ende der Rekrutierungsphase nahmen 15 SE mit 16 Zentren an der Studie teil.

Kurz nach Studienbeginn trat die COVID-19-Pandemie auf, begleitet von Lockdown-Maßnahmen und einer vorübergehenden Aussetzung des MSP. Zu Beginn stagnierte die Rekrutierung damit vollständig aufgrund des Einladungsstopps für das gesamte Screening von > 5 Wochen (27.03.2020 - 30.04.2020). Danach folgte weiterhin im gesamten MSP eine reduzierte Rekrutierung durch Erkrankungen und Quarantänen des Personals, und mit Ängsten bei Screening-TN und Personal im gesamten MSP. Personalabwanderung führte auch in den Folgejahren nach 2020 zu weiteren Problemen in fast allen SE (auch außerhalb dieser Studie) bundesweit.

Eine besondere Herausforderung im gesamten MSP war es, dass die während Screeningstopp, Quarantäne und Personalausfällen nicht durchgeföhrten MX-Screening-Untersuchungen nachzuholen waren. Damit reduzierten sich die Möglichkeiten der SE, für die Studie zu rekrutieren. Neben den Personalengpässen hatten die Studienzentren zusätzlich damit zu kämpfen, die wöchentlich angedachte Fallzahl zu erreichen, denn Personal und die ScrTN

hatten noch lange nach der Pandemie teilweise die Sorge, sich ggf. auch bei Ergänzungsuntersuchungen in der SE anzustecken.

Dies alles reduzierte die Rekrutierung für die Studie derart, dass die ursprünglich geplante Anzahl von Studien-TN trotz dreimaliger kostenneutraler Verlängerungen nicht erreicht werden konnte.

Während der gesamten Studienlaufzeit konnte trotz multipler Schwierigkeiten eine hohe Motivation aller beteiligten Ärzte beobachtet werden. Als Grund für das hohe Engagement der SE in schwierigsten Zeiten wurde immer wieder die Unterstützung des Erkenntnisgewinns für Screening und teilnehmende Frauen durch die wissenschaftliche Studie genannt.

Eine wichtige Maßnahme der Konsortialführung bestand in Motivation der teilnehmenden SE in Form von regelmäßigen Studienverlaufsgesprächen und einer engmaschigen Betreuung der Studienzentren, bei denen wechselseitig über Hürden, Schwierigkeiten aber auch Erfolge berichtet wurde.

Die logistische Herausforderung, in den Screeningablauf die zusätzliche Ultraschalluntersuchung zu integrieren, wurde von den meisten SE auch deshalb bewältigt, da die vorausschauend geplante Studiensoftware den SE die Dokumentation maximal erleichterte, so dass nur der zusätzliche US und die KK separat zu dokumentieren waren.

Alle SE waren sehr bemüht, die geplanten TN-zahlen zu liefern, wenngleich die für die Studie geplanten Fallzahlen reduziert werden mussten, da sie die Kapazitäten der SE innerhalb des Studienzeitraumes überstiegen.

Sehr erfreulich ist, dass trotz der erschwerten Bedingungen die final erreichte Anzahl an Studienteilnehmerinnen ausreichend ist (auch weil der beobachtete Effekt stärker als angenommen war) und damit valide Aussagen zur Effektivität der MXUS getroffen werden können. Dass dieses möglich wurde, ist insbesondere dem hohen Engagement aller teilnehmenden SE zu verdanken.

Zu Beginn war es auch schwierig, ScrTN für die Interviewteilnahme zu gewinnen. Dies lag zum einen daran, dass die Rekrutierung im allgemeinen MSP-Routinebetrieb zunächst "vergessen" wurde, zum anderen war anfangs keine Aufwandsentschädigung für die teilnehmenden Frauen vorgesehen. Nachdem jedoch eine Aufwandsentschädigung von 50 Euro für Interview-TN eingeführt wurde, hatte dies einen positiven Einfluss auf die Rekrutierung und die Teilnahmebereitschaft.

## 2.5 Dichtebestimmung

Die Dichtemessung an sich ließ sich logistisch gut integrieren. Die verwendete Studien-Software informierte die RFA (durch Aufpoppen eines Informations-Fensters) zeitnah nach Abschluss der MX, wann immer der Schwellenwert überschritten war und die TN für die Studie geeignet sein konnte.

Zeitverzögerungen durch die Dichte-Software lagen im Sekundenbereich. Je nach technischer Vernetzung innerhalb der SE konnten in Einzelfällen durch verlängerte Laufzeiten der

fertiggestellten MX in das lokale PACS-System Verzögerungen im Minutenbereich auftreten. Bei guter Vernetzung zum PACS innerhalb der SE gab es keine relevanten Verzögerungen.

Insgesamt erwies sich die software-mäßige Integration der Dichtemessung als sehr praktikable implementierbare Funktion für risikoadaptiertes bevölkerungsbezogenes Screening. Für Mammobile ohne direkte Anbindung an bzw. mit langen Laufzeiten in das zentrale PACS-system der SE wären allerdings abgeänderte Lösungsmöglichkeiten zu suchen.

Die Akzeptanz bei Ärzten und Personal für die gewählte Integration der Dichtemessung war hoch. Für ein flächendeckendes Angebot müssten technische Voraussetzungen in jeder SE geschaffen werden.

Die Dichtesoftware lieferte geräteabhängige Ergebnisse, weswegen eine Anpassung der Schwellenwerte geräteabhängig erfolgen musste. Aufgrund der sehr variablen Verteilung der Dichtewerte in den jeweils vor Ort vorliegenden Populationen, sollte die hiermit erreichbare Genauigkeit durch gezielte ergänzende Studien überprüft und die Schwellenwerte noch optimiert werden. Dieses war nicht Inhalt dieser Studie.

Ein wesentliches Ziel der Studie war es, herauszufinden, inwieweit die beteiligten Ärzte, das Personal und die Studien-TN die zusätzliche Ultraschalluntersuchung im Rahmen des Screeningprogramms für umsetzbar halten. Die Akzeptanz seitens der teilnehmenden Studienzentren (Ärzte und Praxen) war grundsätzlich positiv, aber für die Praxen bedeutete die Studie eine hohe Mehrbelastung, die seitens der Personaldecke nicht immer leistbar war. Das kann man u. a. an der Laufzeitverlängerung der Studie ablesen.

Die Frauen nahmen die Studienteilnahme als Vorteil gegenüber dem nicht risikoadaptierten Screeningprogramm wahr.

## 2.6 Rechtsgrundlage

Die Studien-TN wurden aus dem MSP rekrutiert. Das Programm zur Früherkennung von Brustkrebs wurde deutschlandweit 2004 eingeführt und bis 2009 waren die Strukturen geschaffen, den Frauen zwischen 50 und 69 Jahren flächendeckend alle zwei Jahre eine Screening-MX anzubieten. Dies erfolgt seither mithilfe eines zentralen Einladungswesens, bei dem bundesweit 14 Zentrale Stellen Einladungsschreiben an alle anspruchsberechtigten Frauen versenden.

Die rechtlichen Grundlagen des MSP sind das SGB V (§25a) und die vom G-BA verabschiedete Krebsfrüherkennungs-Richtlinie (KFE-RL). Die Vertragspartner des MSP sind der GKV-Spitzenverband und die Kassenärztliche Bundesvereinigung, die im Bundesmantelvertrag-Ärzte (Anlage 9.2 BMV-Ä) die Organisation des MSP regeln.

Rechtsgrundlage hinsichtlich der Studie innerhalb des MSP war bei vorliegendem positiven Ethikvotum eine unterschriebene Einwilligungserklärung: allen Frauen, die die Einschlusskriterien für DIMASOS2 erfüllten erhielten eine Informationsbroschüre mit Einwilligungserklärung. Eine Wegeversicherung wurde abgeschlossen für die z. T. wiederholten Termine, zu denen Studien-TN erscheinen mussten.

Mit allen Projektbeteiligten wurden Verträge geschlossen, in denen die Rechte und Pflichten geregelt waren.

### **3 Methodik**

Es handelte sich um eine Studie im Mixed-Methods Design mit quantitativen und qualitativen Studienanteilen sowie einem gesundheitsökonomischen Anteil in Form einer Budget-Impact-Analyse (BIA).

#### **3.1 Quantitativer Studienteil**

##### **3.1.1 Studiendesign**

Es handelt sich um eine Kohortenstudie bestehend aus IG und KG. Die Fallzahlplanung lag bei 30.000 TN. Das ursprüngliche Vorhaben, eine historische KG als Vergleichsgruppe zu nutzen, wurde nach der Absprache mit dem G-BA und der Erstellung eines Amendments durch eine parallel rekrutierte KG ersetzt. Dabei wurden diejenigen Frauen, denen keine MXUS angeboten wurde, in die KG eingeschlossen. Es erfolgte bei allen MX der teilnehmenden Standorte eine automatisierte Dichtemessung, anhand welcher eine Studieneignung (sehr hohe Dichte) festgestellt wurde.

Die TN der IG erhielten als Studienmethode MXUS, wohingegen die TN der KG nur MX erhielten.

Es erfolgten umfangreiche Qualitätssicherungsmaßnahmen hinsichtlich der Gerätetechnik und Personalschulung.

Die statistische Auswertung und die dazugehörigen Berichtsteile wurden vom Konsortialpartner erstellt. Dabei wurde die IG mit der KG verglichen sowie intraindividuell die IG (MXUS versus MX) betrachtet.

##### **3.1.2 Studienpopulation**

###### **3.1.2.1 Bildung der Studiengruppe**

Die Studienpopulation bestand aus TN des regulären MSP (Alter 50 bis 69 Jahre), bei denen im MX eine sehr hohe DKD automatisiert gemessen wurde. Die Definition von hoher DKD in dieser Studie entspricht dem oberen Teil von ACR C sowie ACR D. In der Literatur wird unter „hoher Dichte“ häufig der gesamte Dichtebereich der Gruppen ACR C und D zusammengefasst, dieser umfasst ca. 50 % der weiblichen Population zwischen 50-69. Da eine ergänzende Ultraschall-Untersuchung von 50 % der ScrTN keineswegs umsetzbar und auch medizinisch unter Risiko-Nutzen-Aspekten nicht sinnvoll erscheint, war es das Ziel dieser Studie, den Bereich der oberen 15 % gemessenen Dichtewerte zu selektieren.

Für die Dichtebestimmung aus der digitalen MX wurde die Dichte-Software densitasdensity™ der Fa. Densitas (Halifax, Kanada) verwendet, deren Dichtewerte von 1 - 100 skaliert sind. Der Schwellenwert für den Studieneinschluss wurde so definiert, dass die ca. 15 % Frauen mit den

höchsten Densitas-Dichtewerten eingeschlossen wurden. Der Dichtewert wurde hierfür prospektiv anhand von Vorversuchen für die unterschiedlichen MX-Gerätetypen ermittelt.

Potentiellen IG-TN wurde eine Studienteilnahme nur angeboten, wenn diese ihre Screening-MX in den jeweils vordefinierten, den ScrTN nicht bekannt gegebenen Studien-Rekrutierungszeiten erhielten. Außerhalb der Rekrutierungszeiten erfolgte kein Angebot. Die Rekrutierungszeiten wurden variabel, prospektiv und entsprechend der verfügbaren Ressourcen (Personal, Räumlichkeiten) durch die PVÄ der teilnehmenden SE festgelegt. Zudem konnte die Rekrutierung durch das Personal bei Überlastung gestoppt werden. Frauen im ausgewählten Dichtebereich, die kein Angebot für MXUS erhielten, wurden in der lokalen Studien-Software dann erfasst als „kein Angebot“.

Erst nach Berücksichtigung der Ein- und Ausschlusskriterien (siehe 2.2 Beschreibung/Darstellung des Projekts) konnten interessierte Frauen für die IG rekrutiert werden. Von 101.833 TN, die an den von den einzelnen SE vorher bestimmten Studentagen eine Dichtemessung erhalten hatten und oberhalb des Dichtegrenzwertes lagen, wurde 51.087 TN die Studienteilnahme angeboten. 25.746 TN lehnten das Studienangebot ab oder haben keinen (passenden) US-Termin erhalten bzw. diesen nicht wahrgenommen.

25.341 TN wurden in die IG eingeschlossen und erhielten einen ergänzenden US mit MXUS-Befundung.

Der ergänzende US wurde am selben Tag oder binnen der sieben folgenden Kalendertage nach der Screening-MX durchgeführt.

### **3.1.2.2 Bildung der Kontrollgruppe**

Die KG setzte sich aus geeigneten Fällen zusammen (sehr hohe Dichte, kein Ausschlusskriterium), bei denen die MX außerhalb der Rekrutierungszeiten („kein Angebot“) anfiel. Diese Frauen erhielten keinen ergänzenden US.

Die Organisation von Zeiten für Rekrutierung bzw. „kein Angebot“ erfolgte prospektiv durch die/den PVÄ wie oben (siehe 3.1.2.1 Studiengruppe) beschrieben. Diese Strategie ist gegenüber einer historischen Vergleichskohorte von Vorteil, da im Vergleich zu einer historischen Gruppe, Verzerrungen durch den Zeitfaktor vermieden werden konnten. Die Zuteilung von TN zur IG oder KG ergab sich damit weitgehend zufällig durch die Einbettung der Studie in ein Real-World-Setting.

Die anonymisierten Einzelfalldaten dieser Frauen beinhalteten alle für die Auswertung des MX-Screenings erforderlichen nicht-personengebundenen Daten sowie die in der MX gemessenen Dichtewerte.

### **3.1.3 Fallzahlplanung**

Ursprünglich war es geplant, dass die Hypothese (erwartete Steigerung der CA-Entdeckungsrat um 2 pro 1000) mit einer Präzision von + / - 0,5 pro 1000 ermittelt werden soll, entsprechend einem 95 %-Konfidenzintervall von 1,55;2,57. Hierzu wäre eine Fallzahl von 30.000 TN erforderlich gewesen.

Da die geplante Fallzahl von 30.000 TN in der IG nicht erreicht werden konnte, erfolgte eine Anpassung auf 25.000 TN. Trotz der Reduzierung der Fallzahl, die die Power der Studie

beeinträchtigt, konnten zuverlässige Effektschätzer ermittelt werden, allerdings wurde eine Präzisionsminderung von etwa 11 % zum Zeitpunkt der Fallzahlreduktion kalkuliert.

Die finalen Studiengruppen umfassen 25.341 TN in der IG und 38.529 in der KG. Das erreichte 95 %-Konfidenzintervall, lag bei 2,1;3,4 pro 1000.

### **3.1.4 Untersuchungsmethode**

#### **3.1.4.1 Das MXUS-Verfahren**

Die IG erhielt ergänzend zur qualitätsgesicherten Screening-MX einen qualitätsgesicherten US. Die Befundung erfolgte jeweils durch Kombination der Informationen aus beiden bildgebenden Methoden (MXUS). Diese Befundungen erfolgten jeweils durch Befunder\*innen, die sowohl MX als auch US-Erfahrung aufwiesen und zudem eine sonographische Zusatzausbildung erhalten hatten (siehe 3.1.6 Qualitätssicherung und Monitoring).

Die zu testende Studienmethode bestand aus MXUS (nicht der alleinige US), diese Methode wurde mit einer alleinigen MX verglichen. Die Wahl dieser Studienmethode basiert auf der Annahme, dass die Kombination beider Verfahren mehr diagnostische Informationen liefert als jede Methode für sich alleine (mit potentieller Auswirkung auf Sensitivität und Spezifität). Zudem benötigt jede untersuchte Frau nach Durchführung beider Methoden ohnehin eine abschließende Gesamtbefundung und -wertung und keine Einzelbefunde.

Die Schulung der Befunder\*innen beinhaltete, wie MX- und US-Befunde zur Optimierung von Sensitivität und Spezifität kombiniert werden sollten:

- Beispiel 1: Wenn ein US unklarer Befund auffällt, der retrospektiv einem MX-Befund zugeordnet werden kann, der seit Jahren unverändert oder ohnehin typisch gutartig ist, sollte der kombinierte Befund „gutartig“ lauten: Hiermit können unnötige fpos Befunde vermieden werden.
- Beispiel 2: Gerade in großen Brüsten können Schallschatten an interponierten lipomatösen und bindegewebigen Strukturen zu sonographisch verwirrenden Bildern führen. Die Kombination der Information von MX und US erlaubt es, sich bei Durchführung des US auf MX unzureichend beurteilbare Areale oder fraglich auffällige Areale zu konzentrieren, wodurch die Sensitivität des US gesteigert werden kann.

### **Methodische Begriffsbestimmungen**

Die Gruppe von TN mit hoher DKD, welche die kombinierte Screeninguntersuchung mit Mammographie und Ultraschall (MXUS) erhalten haben, wird in der Folge auch als IG bezeichnet. Die Gesamtdiagnosen der IG, die auf MX und der Studienmethode beruhten, wurden mit MXUS bezeichnet. Innerhalb der IG lassen sich die Tumoren abgrenzen, die bereits in der MX sichtbar waren. Diese Tumore werden als MX entdeckt bezeichnet. Die Tumoren in der IG, die nicht durch MX, sondern nur durch den Brustultraschall mit MXUS-Befundung gesehen wurden, werden als US-entdeckt bezeichnet. Hierbei ist zu berücksichtigen, dass es sich nicht um einen alleinigen US handelt.

### **3.1.4.2 Medizinischer Ablauf bis Konsensusbefundung/ DIMASOS2-Screeningkette**

Studien-geeigneten TN wurde die Studienteilnahme angeboten. Die Einladung zur Studienteilnahme erfolgte mit Aushändigung der Studieninformation durch die radiologischen Fachkräfte (RFA) unmittelbar nach der Screening-MX. Bei Bedarf konnte die TN auch ein ärztliches Gespräch erhalten. Bei Interesse wurde ein US-Termin am selben Tag oder innerhalb der sieben folgenden Kalendertage angeboten.

Geeignete TN wurden erst dann in die IG eingeschlossen, wenn sie die Einverständnis- und Datenschutzerklärung unterschrieben hatten und zum US-Termin erschienen waren.

Damit ergaben sich für geeignete TN folgende Status-möglichkeiten:

- Kein Angebot → KG
- Ausschlusskriterium (Implantat, Teilnahme an einer anderen diagnostischen Studie, fehlende Einwilligungsfähigkeit) lag vor
- Ablehnung (durch ScrTN)
- Teilnahme → IG

TN der IG erhielten mit Einschluss in die Studie einen zusätzlichen US. Vor Durchführung des US sollte der US-Untersucher obligat zuerst die MX analysieren und, wenn möglich (s. unten), als MX-Befund in der Software dokumentieren. MX-EB und MX-ZB waren (auch entsprechend Screening-Gesetzgebung) immer gegeneinander verblindet. Die Pflicht zur Dokumentation des MX-Befundes durch den sonographierenden Arzt/Ärztin entfiel nur dann, wenn der MX-EB und der MX-ZB bereits (durch ein anderes MX-Befunderpaar) dokumentiert und damit abgeschlossen waren. Nach Durchführung des US erstellte der/die US-Befunder\*in dann seinen/ihren MXUS-EB.

Die mammographische Zweitbefundung (MX-ZB) erfolgte ebenfalls systematisch ohne Kenntnis der US-Untersuchung. D. h., zu jedem Fall der IG war zuerst ein MX-ZB zu erstellen. Danach konnte durch denselben oder einen anderen Befunder\*in durch Aufruf der ergänzenden US-Bilddokumentation im Patientenarchiv (PACS) der MXUS-ZB erstellt werden.

EB und ZB erfolgten für MX und für MXUS-Befunde jeweils unabhängig voneinander. Diese Verbindung aller EB und ZB war auch in der Logistik der Studiensoftware (für die Befund-Dokumentation) sichergestellt.

Bei positiven EB und/oder ZB folgte immer eine Konsensuskonferenz (KK), die vom/von der PVÄ geleitet wurde. Generell wurde die MX-KK vor der MXUS-KK durchgeführt. Bei der KK waren üblicherweise (entsprechend gesetzlichen Screening-Vorgaben) die MX-Befunder und MXUS-Befunder anwesend.

### **3.1.4.3 Befundungsergebnisse in der Interventionsgruppe**

Ein Befund (analoges Vorgehen bei MX und MXUS, entsprechend dem im MSP vorgeschriebenen Prozedere) wurde als unauffällig/benigne (negativ) gezählt, wenn:

- EB und ZB unauffällig/benigne befundet waren oder
- bei auffälligem EB und/oder ZB, die KK unauffällig/benigne war.

Ein Befund (analoges Vorgehen bei MX und MXUS) wurde als auffällig/evtl. maligne (positiv) gezählt:

- Bei auffälligem KK-Befund.

Damit ergab sich für jeden Fall ein positiver oder negativer MX-Endbefund und ein positiver oder negativer MXUS-Endbefund.

Normalerweise würde bei alleiniger MX-Befundung ein positiver MX-Endbefund eine BA auslösen; ein negativer MX-Endbefund würde keine BA auslösen.

Bei MXUS-Befundung würde ein positiver MXUS-Endbefund eine BA auslösen; ein negativer MXUS-Endbefund würde keine BA auslösen.

Da die oben genannten Konstellationen dem Real-World-Setting entsprechen, gingen die oben genannten Endbefunde auch so in die Studienauswertung ein.

Gemäß gesetzlichen Vorgaben war allerdings gefordert, dass immer, wenn ein positives MX-Ergebnis oder ein positives MXUS-Ergebnis vorlag (unabhängig vom Ergebnis der jeweils anderen Studienmethode), eine Abklärung (BA) erfolgen musste. Daher ergaben sich etwas mehr BA, als durch die jeweils betrachtete Studienmethode (MX oder MXUS) erforderlich gewesen wäre.

Für jede Methode wurde jeweils fallbezogen und nicht brust- oder läsionsbezogen ausgewertet.

#### **3.1.4.4 Nomenklatur**

Die Studienmethode war ein ergänzend durchgeführter Ultraschall, der aber immer zusammen mit den Informationen aus der MX bewertet wurde. Diese Methode ist trotz der kombinierten Befundung von MX und US (MXUS) in den Tabellen und statistischen Auswertungen vereinfacht als „US“ bezeichnet.

Beim intraindividuellen Vergleich innerhalb der IG wurden folgende Bezeichnungen verwendet:

„MX“	= Screeningbefund, der nur auf der Screeningmammographie beruht, also alle Befunde vor Durchführung der Testmethode
„US“	= alle Befunde, die <u>nur</u> auf der Testmethode beruhen, also nicht bereits in der Screeningmammographie gesehen wurden.
„Insgesamt“	
oder „MXUS“	= alle Befunde der IG nach Durchführung der Testmethode. Sie beinhalten also alle Befunde, die auf „MX“ beruhen und alle, die nur auf der Testmethode „US“ beruhen.

#### **3.1.4.5 Medizinischer Ablauf ab bildgebender Abklärung**

Die BA erfolgte bei allen Frauen mit Indikation zur weiteren Klärung (siehe 3.1.4.3 Befundungsergebnisse in der Interventionsgruppe) nach internationalen Standards und entsprechend den Vorgaben des MSP durch die/den PVÄ.

Für die Abklärungsuntersuchung wurden diese Frauen nach Abschluss der KK zu einem separaten Termin eingeladen.

Im Rahmen der BA konnte sich ein negativer/gutartiger Befund ergeben (entspricht BI-RADS 2) mit Empfehlung zur weiteren üblichen Überwachung im MSP (alle 2 Jahre). Bei vermutlich gutartigen Befunden (BI-RADS 3) wurde eine kurzfristige Kontrolle empfohlen.

Bei positivem Ergebnis der BA wurde eine histologische Abklärung empfohlen. Üblicherweise erfolgte diese als sonographisch-gestützte Stanzbiopsie, als stereotaktisch-gestützte Vakuumbiopsie bei nur mammographisch sichtbaren Befunden und nur sehr selten eine operative Abklärung. Eine operative Abklärung (offene Biopsie) wurde dann empfohlen, wenn eine Patientin einer minimal invasiven Untersuchung nicht zustimmte oder sich (äußerst selten) nicht dafür eignete. Die histologischen Ergebnisse wurden in multidisziplinären Fallkonferenzen in den SE gemäß dem internationalen Standard besprochen. Hierbei wurde auch eine Empfehlung zur weiteren Behandlung (z. B. bei gesicherten Malignomen Operation und/oder Chemotherapie; bzw. bei B3-Läsionen, wie z.B. atypischer duktaler Hyperplasie (ADH), FEA oder vergleichbaren Läsionen, entweder Rebiopsie oder Kontrolle) ausgesprochen.

### **3.1.4.6 Dokumentation der Daten**

Für die prospektive Dokumentation aller Studienfälle wurde durch die Fa. IdeeDesign Michael Hertlein in Zusammenarbeit mit der Konsortialführung eine Online-Studien-Software entwickelt.

Sie ermöglichte die Übertragung aller für die Befundung relevanten Daten aus den offiziellen Softwares in die Studiendatenbank, die Planung von Untersuchungszeiten und die kontinuierliche Dichtemessung (durch Integration der kommerziell verfügbaren Dichtesoftware). Bei Überschreitung des Schwellenwertes löste die Studiensoftware die Einladung von Frauen (je nach standortspezifischer Organisation: durch die radiologische Fachkraft oder Anmeldung) aus. In der Software wurde die Teilnahme/Nicht-Teilnahme bzw. Ablehnung oder fehlende Eignung (in KG oder bei Ablehnung bzw. fehlender Eignung anonymisiert) dokumentiert. Die Studiensoftware unterstützte die komplette schrittweise Erfassung der Studienbefundung und garantierte eine gesetzeskonforme Befunderkommunikation (inklusive Befund-Verblindung zwischen EB und ZB – je nach Untersuchungsschritt) im Rahmen der Studie. Dieses Vorgehen war für die Studienmethode analog dem für MX vorgeschriebenen Vorgehen im MSP. Weiterhin ermöglichte die Studiensoftware die Online-Rückübertragung der Studienergebnisse in die offiziellen Softwares entsprechend Studienprotokoll und gesetzlichen Vorgaben sowie die gesetzlich geforderte Initiierung weiterer BA, wo erforderlich (bei positivem MX-Endergebnis und/oder positivem MXUS-Endergebnis). Die in den SE durchgeführten BA und weiteren Schritte waren (wie im MSP üblich) durch die SE direkt in die offizielle Software einzupflegen. Auch die Zusammenführung der Ergebnisse weiterer BA bis inklusive der postoperativen Konferenz (durch quartalsweises Abgleichen von Masterdatenexporten) mit den Einzelfalldaten war eingerichtet. Die Software unterstützte zudem das Monitoring der teilnehmenden SE, die Auszahlung der Studien-Erstattungen an die SE sowie den regelmäßigen anonymisierten Datenexport an das ISE für deren statistische Zwischen- und Endauswertungen.

Die gesamte Dokumentation der Studienfälle, d. h. alle relevanten Daten aus den Screeningsoftwares und aus der Studie, wurden in einem ausschließlich lokal in jeder SE vorliegenden Anteil der Studiensoftware gespeichert. Hieraus wurden die für die Studie relevanten Einzelfalldaten datenschutzgerecht in die zentrale Studiendatenbank übertragen.

Die Dokumentation und Komplettierung der Einzelfiles der KG erfolgte analog zur IG in den hier allerdings anonymisierten Einzelfiles.

### **3.1.5 Automatisierte Dichtebestimmung**

Ziel der Studie war es, den oberen Dichtebereich (ca. oberen 15 % aller in der Population vorkommenden Dichtewerte) in die Studie einzuschließen. Dieser Bereich entspricht ACR D, welches die oberen 5 % der Dichtewerte in der Population einschließt und einem oberen Dichtebereich von ACR C, also weitere ca. 10 % des Anteils hoher Dichtewerte in der Population. Die Gesamtgruppe aller ACR C-Fälle hingegen beträfe ca. 40 - 45 % der Population (Weigel 2020; Wanders 2019).

Die Dichtemessung erfolgte mit der kommerziell verfügbaren CE-zugelassenen automatisierten Dichtesoftware *densitasdensity™* der Fa. Densitas (Halifax, Kanada). Diese bestimmt die Dichte mithilfe von KI aus volldigitalen prozessierten Bilddaten, wie sie von jedem MX-Gerät im sogenannten DI-COM-Standard geliefert werden.

Das Programm war für die häufigsten MX-Gerätetypen zugelassen. Im Rahmen der Vorbereitungsphase des Projekts wurden weitere Gerätetypen ergänzt und eine Dichteskaliierung von 1 - 100 (mithilfe von KI) durch die Firma ermöglicht. Die so skalierten Dichtewerte werden als Densitas-Dichte-Score bezeichnet. In Vorversuchen war anhand von 17.627 Messwerten aus unterschiedlichen Regionen für die in der Studie verfügbaren Gerätetypen ein Schwellenwert so gewählt, dass ca. die oberen 15 % der Population eingeschlossen werden konnten.

### **3.1.6 Qualitätssicherung und Monitoring**

Die Qualitätssicherung umfasste eine Eingangsprüfung der US-Gerätetechnik, wobei nur Geräte der oberen Mittelklasse und Oberklasse zugelassen waren. Eine Eingangsüberprüfung der Schallköpfe erfolgte durch die Studienleitung und danach regelmäßig durch die PVÄ.

Alle in der Studie tätigen verantwortlichen Befunder\*innen waren zugelassene Screeningbefunder\*innen mit ausreichender MX-Vorerfahrung. Sie mussten zudem ausreichende Erfahrung im US nachweisen (Spezialausbildung und Tätigkeit als PVÄ oder Vertretungs-PVÄ oder vorliegende Kassenzulassung bzw. bei Krankenhausärzt\*innen mindestens DEGUM 1-Zertifikat sowie eine Vorerfahrung > 1000 sonographischen Untersuchungen in 3 Jahren).

Vor Start musste jede\*r MXUS-Befunder\*in einen zusätzlichen Kurs für MXUS ableisten, der von der Studienleitung durchgeführt wurde.

Während der gesamten Rekrutierungsphase erfolgte durch die Konsortialführung ein kontinuierliches Monitoring mit Stichprobenprüfungen, Plausibilitätschecks der Datenexporte und Nachschulungen bei Neueinstieg von Personal in den SE.

Die Ergebnisse wurden auch als Benchmarks den SE regelmäßig mitgeteilt. Sie wurden mit jeder SE in regelmäßigen Studienverlaufsgesprächen zwischen der Studienleitung und den Studienzentren diskutiert.

Nach Abschluss der Rekrutierung wurden aus der Studiensoftware sämtliche Daten exportiert, geprüft, mit der SE von Fehlern bereinigt und pseudonymisiertes Bildmaterial von den CA-Fällen für die Studienleitung verfügbar gemacht.

### **3.1.7 Statistische Auswertung**

Die im Rahmen von DIMASOS2 gewonnen anonymisierten Daten der Studiendatenbank wurden von der Studienleitung zwei Mal pro Jahr über eine institutseigene Nextcloud anonymisiert an das ISE übermittelt.

Für die Auswertungen wurde frühzeitig ein ausführlicher statistischer Analyseplan erstellt (15.01.2020), weiterentwickelt und dem Innovationsausschuss mitgeteilt. Die in diesem Bericht vorgelegten Auswertungsergebnisse orientieren sich an diesem Analyseplan (siehe 3.1.7.3 Auswertung der Daten/ Statistische Methoden).

Die Auswertungssyntax wurde anhand der regelmäßigen Datenauszüge ebenfalls frühzeitig entwickelt und weiter ausdifferenziert.

Es wurden regelmäßig Zwischenauswertungen zur Anzahl und Klassifikation der zusätzlich gefundenen CA sowie des Einflusses auf die diagnostische Genauigkeit des MSP durchgeführt.

Die Auswertung erfolgte für folgende TN-Gruppen:

- KG: Hierunter fielen alle Frauen mit Eignung und ohne Ausschlusskriterien, deren Screening-MX außerhalb der Rekrutierungszeit stattfand. Diese TN erhielten keinen Ultraschall.
- IG: Hierunter fielen alle TN mit Eignung und ohne Ausschlusskriterien, die nach entsprechender Einwilligung eine US-Untersuchung erhielten und deren Diagnose sich aus der mammographisch-sonographischen kombinierten Untersuchung ergab.

Innerhalb der IG wurden die Befundungen, wie folgt (vergleiche 3.1.4.4 Nomenklatur), unterteilt.

#### **3.1.7.1 Datenaufbereitung bei der Studienleitung**

Die Datenaufbereitung war ein iterativer Prozess mit folgenden Komponenten:

##### **1. Prüfung auf vollständige Dokumentationen**

Hierfür wurden anhand von Zwischenexporten umfangreiche Plausibilitätsprüfungen durchgeführt. Es wurde insbesondere auf Dokumentationslücken (fehlende Werte, z. B. fehlende Angaben von CA-Stadien oder CA-Biologie) und Fehldokumentationen (unplausible Angaben, z. B. Diskrepanz von Größenangabe und Stadium, Seitenverwechslungen etc.) gecheckt. Diese wurden zusammen mit den SE bereinigt, ggf. nach ergänzenden Recherchen durch die SE.

##### **2. Prüfung des Exports aus der Screening-Software bezüglich logischer Konsistenz**

Nach dem Export der Datenbank wurden logische Lücken im medizinischen Ablauf (unklare Zuordnung zwischen Einträgen in unterschiedlichen Tabellen zueinander) und Fehldokumentationen (Mehrfach- und Beidseits-Befunde mit z. T. wechselnden

Läsionsnummern) ermittelt. Diese wurden detektiert und inhaltlich recherchiert (Arztbriefe, Entlassungsbriebe, Bildmaterial). Dieser Prozess musste für konsistente Daten mehrfach wiederholt werden.

### 3. Korrekte Reduktion der Datenbankstruktur auf eine Auswertungstabelle

Für die statistische Auswertung war eine tabellarische Aufbereitung der Fälle notwendig. Die Reduktion der komplexen Datenbankstruktur führte zum Teil zu fehlerhaften Vereinfachungen oder zeigte Inkonsistenzen in den Daten auf.

Bei Fällen mit Mehrfachbefunden wurde anhand der Dokumentation der schwerwiegendste Befund ermittelt, da definitionsgemäß die Auswertung Patienten-bezogen, nicht Seiten-bezogen oder Befund-bezogen erfolgen sollte. Nach Recherchen händisch nachgetragene Daten wurden nach erneuten Exporten in neuen anonymisierten Datenbankabzügen zur Auswertung dem Konsortialpartner bereitgestellt.

Bei Mehrfachbefunden (sowie bei eventuell zusätzlich bestehenden Lücken oder fehlerhaften Dokumentationen) bedingte dieses zeitintensive Recherchen, Rücksprachen mit den teilnehmenden PVÄ und Plausibilitätschecks in einem mehrstufigen, wiederholenden Vorgehen.

#### 3.1.7.2 Fragestellung

(1) *Primäre Hypothese* war, dass durch die Studienmethode ergänzend zum MSP > 2 / 1000 CA mehr entdeckt werden.

Hierfür war zu bestimmen: die Karzinomdetektionsrate für die kombinierte Untersuchung (MXUS) verglichen zur Detektionsrate des MSP mit alleiniger MX.

Parallel dazu waren folgende Fragen zu beantworten:

(2) Beschreibung der zusätzlich gefunden CA (Klassifikation, Stadien)

(3) Bestimmung der fpos Befunde (Wiedereinbestellungsrate, BX-Rate, 6-Montas-Kontroll-Empfehlungen)

(4) Bestimmung der Sensitivität des MSP für die kombinierte Untersuchung (MXUS)

(5) BIA zur Modellierung des Kosten-Nutzen Verhältnisses

(6) Befragung (Telefoninterviews) zur Machbarkeit und Relevanz des zusätzlichen US unter Verwendung von semi-strukturierten Interviewleitfäden, durchgeführt mit (a) TN der Untersuchung sowie (b) den beteiligten Fachkräften (ärztliches Personal und RFA).

#### 3.1.7.3 Auswertung der Daten/ Statistische Methoden

Für die Auswertungen wurden die Daten zunächst mit üblichen statistischen Lage- und Streuungsmaßen deskriptiv, auch in Untergruppen beschrieben. Wenn möglich wurden 95 %-Konfidenzintervalle bestimmt.

Für die Überprüfung der primären Hypothese wurde die Zahl der ausschließlich durch die Testmethode diagnostizierten CA bestimmt und die Rate pro 1000 Frauen inklusive Konfidenzintervall ermittelt. Erwartet wurde eine Rate von zusätzlichen 2 pro 1000 (zu den regulär nur durch MX diagnostizierten CA) mit einem 95 %-Konfidenzintervall von 1,55 - 2,57.

Um strukturelle Unterschiede (wie Alter, Erst- / Folgeuntersuchung) zwischen IG und KG auszugleichen wurde eine inverse Propensity Score Gewichtung ("inverse propensity treatment weighting", IPTW) angewendet. Dafür wurde zunächst für jede TN die Wahrscheinlichkeit („Propensity Score“) berechnet, mit der sie der jeweiligen Gruppe zugewiesen wurde. Die Berechnung basierte auf den Kovariablen Alter, Region, Dichte und Screeningrunde. Ausgehend von diesen Wahrscheinlichkeiten wurde für jede TN ein Gewichtungsfaktor berechnet, der umgekehrt proportional zur individuellen Zuweisungswahrscheinlichkeit ist. Durch diese Gewichtung wird eine pseudorandomisierte Population erzeugt, in der die Verteilung der beschriebenen Kovariablen zwischen den Gruppen ausgeglichen ist. Die Fallzahl dieser Population wird durch die Gewichtung künstlich aufgebläht, weshalb eine direkte Interpretation nicht mehr zulässig ist. Die Berechnung von Raten, wie beispielsweise Inzidenzen oder prozentuale Anteile ist jedoch auf Basis der gewichteten Daten möglich, sodass die nun adjustierten CA-Entdeckungsraten der beiden Gruppen unmittelbar miteinander verglichen werden können.

Weiter wurden absolute und relative Differenzen berechnet. Weitere statistische Tests wurden, auch zur Vermeidung von multiplem Testen, nicht durchgeführt. Die statistische Bewertung der Ergebnisse erfolgt, dort wo angezeigt, jeweils auf Basis der ermittelten Konfidenzintervalle.

Tabelle 2 beschreibt die im Analysenplan für die Auswertung definierten Parameter:

*Tabelle 2: Beschreibung der Endpunkte (Outcomes) und die Art der jeweiligen Erfassung*

<b>Outcomes</b>	<b>Beschreibung</b>	<b>Untersuchungsmethode</b>
Tumorentdeckungsrate des zusätzlichen Ultraschalls (US)	Ermittlung der durch den zusätzlichen Ultraschall zusätzlich detektierten Tumoren	Angabe der in der Studiendatenbank dokumentierte Anzahl der gefunden Tumoren, deren Befunde auf der kombinierten Untersuchung beruhen. Es werden Raten pro 1000 TN berichtet.
Tumorentdeckungsrate in der IG (MXUS) und KG (MX)	Ermittlung der in der IG (durch MX und US) und KG (im MSP mit alleiniger MX) detektierten Tumoren	Die in der Studiendatenbank dokumentierte Anzahl der gefunden Tumoren in beiden Gruppen werden nach Anwendung vom Inverse Propensity Treatment Weighting (IPTW) miteinander verglichen. Neben der Angabe von Raten pro 1000 TN werden absolute und relative Differenzen berichtet.
Beschreibung der zusätzlich gefunden Tumoren (US)	Bestimmung der Tumorcharakteristika (Klassifikation, Stadien, Grading, OP-Vorgehen)	Die in der Studiendatenbank dokumentierten Tumoren werden hinsichtlich mehrerer Charakteristika beschrieben (deskriptive Beschreibung).
Bestimmung der falsch-positiv Raten	Gruppenvergleiche zur Wiedereinbestellungsrate, zur	Die in der Studiendatenbank dokumentierten Raten zur

	Rate weiterer durchgeföhrter Untersuchungen sowie zur Biopsierate.	Wiedereinbestellung, zusätzlichen Untersuchungen und Biopsien werden miteinander verglichen. Neben der absoluten Anzahl werden Raten pro 1000 TN berichtet.
Bestimmung der Sensitivität des MSP für die kombinierte Untersuchung (MXUS)	Die Sensitivität des kombinierten Verfahrens wird im Vergleich zum MX für die TN mit dichtem Brustgewebe geschätzt.	Es werden Vierfeldertafeln mit unterschiedlichen Szenarien zur Tumorentdeckungsrate für TN mit dichtem Brustgewebe erstellt. Dabei wird unter der Annahme verschiedener Sensitivitätsangaben der alleinigen MX aus der Literatur und der beobachteten Tumorentdeckungsraten die Sensitivitätssteigerung durch eine kombinierte Untersuchung berechnet.
Bestimmung des Nutzen-Schaden Verhältnisses der kombinierten Untersuchung (MXUS)	Bestimmung der zusätzlich entdeckten Tumoren (Nutzen) im Verhältnis zu den zusätzlich auftretenden BA bzw. BX (Schaden)	Die in der Studienbank dokumentierten Tumoren sowie die BA und BX werden in beiden Gruppen verglichen und ins Verhältnis gesetzt.
Bestimmung der durch den zusätzlichen Ultraschall entstehenden Kosten	Zur Abschätzung der zusätzlich zu erwartenden Kosten für das kombinierte Verfahren wurde eine Budget-Impact-Analyse durchgeführt.	Modellierung eines Markov-Models mit Entscheidungsbaum auf der Basis aktueller GKV-Daten.
Untersuchung der Akzeptanz und Praktikabilität des kombinierten Verfahrens.	Befragung (Telefoninterviews) zur Akzeptanz und Machbarkeit des zusätzlichen US unter Verwendung von semi-strukturierten Interviewleitfäden, durchgeführt mit (a) TN des MSP sowie (b) den beteiligten Fachkräften (ärztliches Personal und RFA).	Qualitative Inhaltsanalyse nach Philipp Mayring.

### 3.2 Gesundheitsökonomische Analyse

Im Rahmen der gesundheitsökonomischen Analyse wurde ein Markov-Modell erstellt, in dem ein Entscheidungsbaum verschachtelt ist, um die ökonomischen Auswirkungen der Intervention zu modellieren. Der Entscheidungsbaum wurde dazu verwendet, die herkömmlichen Kosten des MSP mit den zusätzlichen Kosten für den US und damit verbundene potentielle Folgeuntersuchungen (z. B. BA, BX) sowie den zu erwartenden gesundheitlichen Nutzen in Relation zu setzen. Zum erwarteten Nutzen der kombinierten

Untersuchung zählen: frühere Diagnosen und niedrigere CA-stadien sowie damit verbundene bessere Prognosen, die sich langfristig als geringere ökonomische Belastung auswirken sollten.

Es wurde eine BIA für MXUS erstellt. Es wurden dabei nur direkte Kosten (aus Sicht der GKV) betrachtet (Sullivan 2014). Der relevante Zeitraum für den Kostenträger ist die Zeit zwischen zwei Screening-Runden. In dieser Frist treten alle relevanten Kostenveränderungen durch das zusätzliche Screening auf. Das Ziel dabei war es, die potentiellen zusätzlichen Kosten für die GKV verursacht durch den ergänzenden US für Frauen mit hoher DKD zu bestimmen.

Für die Modellierung gab es zwei Strategien: (1) Frauen, die einen zusätzlichen US bekamen und (2) Frauen der KG, die keinen zusätzlichen US erhielten. Die Datengrundlage bildet die Ergebnisse des quantitativen Studienteils sowie Angaben aus der Literatur ab.

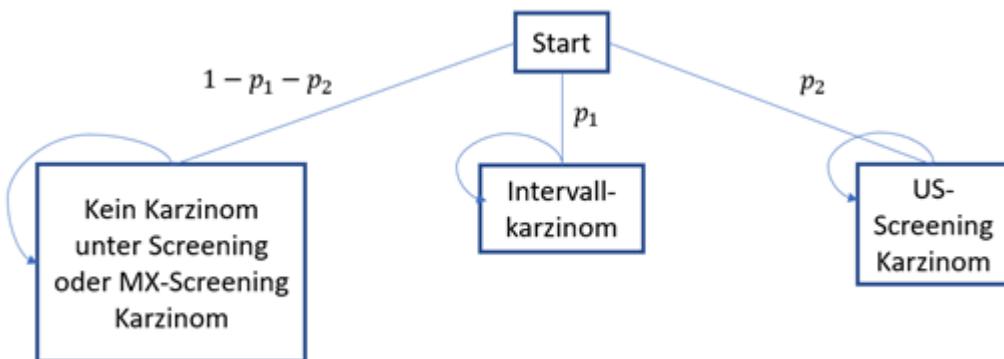


Abbildung 1: Markov-Modell des Budget-Impact Modells mit Übergangswahrscheinlichkeiten  $p_1$  und  $p_2$  („US“: ergänzende US-Untersuchung mit MXUS-Befundung = Studienmethode)

Durch den zusätzlichen US entstehen, neben den Kosten für die US-Untersuchung selbst, weitere Kosten, z. B. durch Abklärungsmaßnahmen und BX bei vielen gutartigen Befunden und bei der begrenzten Zahl zusätzlich entdeckter Malignome. Andererseits kann angenommen werden, dass CA, die zu einem späteren Zeitpunkt als Intervallkarzinome diagnostiziert worden wären, durch den additiven US zu einem prognostisch günstigeren, da früheren, Zeitpunkt, erkannt werden können. CA, die zu einem früheren, klinisch günstigeren Zeitpunkt durch Screening diagnostiziert werden, verursachen in der Regel geringere Behandlungskosten (Khan 2023). Dieser mögliche Effekt wird sich mutmaßlich aber erst nach längerer Zeit einstellen, der von der hier durchgeföhrte BIA nicht abgedeckt wird.

Die BIA wurde über ein Markov-Modell mit Entscheidungsbaum operationalisiert. Ein Entscheidungsbaum ist eine grafische Darstellung von Entscheidungsprozessen, bei der jeder Knoten des Baums eine Entscheidung oder ein Ereignis darstellt und die Äste die möglichen Ergebnisse oder Entscheidungen widerspiegeln. Das Modell besteht aus vier Markov-Zuständen, „Start“, „nur durch die Studienmethode detektiertes Karzinom“, „Intervallkarzinom“ und „kein Karzinom oder Karzinom im MX-Screening“ (siehe Abbildung 1). Es berechnet den Erwartungswert der Kosten pro Person in jeder Strategie. Die Differenz der beiden Erwartungswerte sind die zusätzlichen entstehenden Kosten. Zu der Strategie (1) der zusätzlichen US-Untersuchung gehören alle zusätzlichen Abklärungsuntersuchungen und BX, die dann zum Teil in den zusätzlich detektierten CA resultieren.

Frauen, die eine zusätzliche „US“ - Untersuchung erhalten haben, können in dem Markov-Modell vier potentielle Endpunkte erreichen: (1) Karzinom nur durch „US“ detektiert, (2) Intervallkarzinom, (3) Karzinom detektiert durch MX, (4) kein Karzinom. Die letzten beiden Fälle können zusammengefasst werden, da es in diesem Fall keine Unterschiede zwischen den Kosten der beiden Strategien gibt. Bei den Frauen der KG ist ein im Screening nur durch „US“ detektiertes Karzinom nicht möglich.

Die Übergangswahrscheinlichkeiten von dem Markov-Zustand „Start“ zum Zustand „Intervallkarzinom“ zum Zustand „nur MXUS-Screening Karzinom“ und zum Zustand „Kein Karzinom unter Screening oder MX Karzinom“  $1 - p_1 - p_2$  (Abbildung 1) wurden mit den Szenarien der quantitativen Analysen zur Sensitivität der MX für beide Strategien berechnet. Die Wahrscheinlichkeiten für BA, BX und Pathologien sowie die Verteilung der durch nur MXUS detektierten CA bilden sich aus den Ergebnissen unseres quantitativen Ergebnisteils. Für die Verteilung der Intervallkarzinome wurden die Veröffentlichungen von Armbinder und Niraula et al. (2020) verwendet.

Die Kosten für den zusätzlichen US, BA, BX und Pathologie wurden mit Unterstützung von Expertinnen der Konsortialführung aus den relevanten Ziffern des EBM Katalogs bestimmt (KBV Kassenärztliche Bundesvereinigung, 2024). Die Kosten für die Behandlung in Abhängigkeit des UICC-Stadium des Karzinoms wurden aus der Literatur entnommen (Khan 2023). Aufgrund des Zeitraumes von zwei Jahren für die Modellierung dieser BIA wurde keine Diskontierung vorgenommen (Sullivan 2014). Es wurden die Kosten für das Jahr 2024 betrachtet.

In einer deterministischen Sensitivitätsanalyse wurde mithilfe der 95 %-Konfidenzintervalle überprüft, wie sensitiv das Modell auf Veränderungen der Inputparameter aus den Studienergebnissen reagiert. Zudem wurden mehrere Szenarioanalysen durchgeführt. Diese bezogen sich zum einen auf die verschiedenen Szenarien zur Sensitivität des MSP und zum anderen auf den Anteil von Frauen mit hoher DKD, die US erhalten sollen. Hierbei wurden die oberen 20 %, 15 %, 10 % und 5 % gewählt. Alle gesundheitsökonomischen Analysen wurden mit R Version 4.1.3 durchgeführt (R Core Team, 2019).

Das erste gesundheitsökonomische Modell war Ende 2020 fertiggestellt, das finale Modell lag mit Ende der Rekrutierung im März 2024 vor.

### 3.3 Qualitativer Studienteil

Für den qualitativen Studienteil wurde auf der Basis einer umfassenden Literaturrecherche je ein Interviewleitfaden für TN und einer für medizinisches Fachpersonal entworfen. Mitarbeitende des ISE führten die Telefoninterviews durch und werteten sie inhaltsanalytisch aus.

Zwischen 2020 und 2021 wurden an sechs SE qualitative Telefoninterviews mit medizinischen Fachkräften ( $N = 9$ ) und Frauen mit hoher DKD ( $N = 22$ ) durchgeführt. Die Interviewteilnahme war freiwillig und fand nach schriftlicher Einwilligung der Teilnehmenden telefonisch statt. Alle Interviews folgten einem halbstrukturierten Interviewleitfaden.

Die TN waren zwischen 50 und 69 Jahre alt. Die Rekrutierung erfolgte durch die RFA nach erfolgter Messung der Brustdichte. Jede Frau erhielt eine Broschüre mit Informationen über dichtes Brustgewebe und deren Auswirkungen auf das MX-Screening sowie weitere Hintergrundinformationen zur Studienteilnahme. Die Interviews mit 7 Ärzten (Gynäkologen oder Radiologen), darunter 4 Leiter der teilnehmenden Screening-Zentren, und 2 Befunder\*innen wurden durch das Referenzzentrum München organisiert und von Mitarbeiterinnen des ISE durchgeführt. Dabei nahm das Referenzzentrum Kontakt zu den SE auf und bat das dort beschäftigte Personal, an einem Interview teilzunehmen. Personen, die zu einer Teilnahme bereit waren, kontaktierten zur Terminvereinbarung das ISE. Die angestrebte Anzahl von Interviewpartnern aus dem Bereich Screeningpersonal wurde sehr schnell erreicht. Die teilnehmenden Personen waren Ärzt\*innen (PVÄ und nicht PVÄ) sowie RFA aus vier teilnehmenden SE. Von den Ärzt\*innen waren 3 Personen männlich und 4 Personen weiblich, beide teilnehmenden RFA waren weiblich.

Als deutlich schwieriger gestaltete sich die Rekrutierung von Screening-Teilnehmerinnen in den 6 teilnehmenden SE. Durch starke zeitliche Belastungen des Personals in den SE lief die Rekrutierung anfangs sehr schleppend und eine gezielte Ansprache von geeigneten TN wurde im laufenden Screeningbetrieb häufig vergessen. Zu Beginn war die Zahlung einer Aufwandsentschädigung für die Teilnahme an einem Interview nicht vorgesehen. Da die geplante Anzahl von Interviewteilnehmerin auf diese Weise nicht erreicht werden konnte, wurde die Strategie umgestellt. Durch den finanziellen Anreiz konnte, wie erhofft, eine höhere Teilnahmebereitschaft erzielt werden. Dabei war die Bereitschaft zur Teilnahme das Hauptkriterium für den Einschluss in die qualitative Studie, zudem konnte hinsichtlich des Alters eine ausreichende Diversität erzielt werden.

Die Interviews mit den Frauen begannen mit der Frage, ob sie zum ersten Mal davon gehört hatten, dass sie eine sehr hohe DKD haben und welche Gedanken und Gefühle sie dazu hatten. Anschließend wurden sie zu ihren Erfahrungen mit der zusätzlichen US-Untersuchung befragt. Zudem wurden sie zu ihrer Einschätzung der Wirksamkeit des MSP befragt und ob sich ihre Meinung aufgrund der Informationen über die DKD geändert hätte.

Der Interviewleitfaden für Ärzt\*innen und RFA befasste sich mit deren Wahrnehmung hinsichtlich des Bewusstseins der Frauen für die Brustdichte, den damit verbundenen Auswirkungen auf die Sensitivität des MSP und der Relevanz des Themas für das MSP. Zudem wurden die Durchführbarkeit und Praktikabilität des zusätzlichen US eingeschätzt, einschließlich der Vor- und Nachteile des kombinierten Verfahrens.

Alle Interviews wurden aufgezeichnet und wortwörtlich transkribiert, wobei Orte und Namen unkenntlich gemacht wurden. Die Datenanalyse erfolgte mittels qualitativer Inhaltsanalyse (Mayring P. 2022) unter Verwendung der frei zugänglichen Auswertungssoftware QCMap (Qualitative Content Analysis Program). Zunächst wurden die Interviewdaten unabhängig voneinander von zwei Mitarbeiterinnen des ISE analysiert. Aufgrund des explorativen Ansatzes dieser Studie wurde das Material kodiert, indem induktiv Kategorien aus den Interviewdaten extrahiert wurden.

Um Kodierschemata für (a) die Daten der Fachkräfte und (b) die Daten der TN zu erstellen, begannen die auswertenden Personen mit je zwei Interviews aus jedem Datensatz. Bei der Kodierung wurden ausgewählten Textfragmenten mit ähnlichem Inhalt Codes zugewiesen.

Anschließend wurden durch die Kombination kodierter Fragmente mit ähnlichem Inhalt Themen identifiziert, die wiederum zu übergreifenden Themen zusammengefasst wurden. Im Rahmen von regelmäßigen Treffen wurden Themen, Kodierungen und Interpretationen diskutiert, wobei sich schließlich auf je ein Codeschema für Screeningpersonal und eines für TN geeinigt wurde. Die verbleibenden Interviews wurden anschließend anhand dieser entwickelten Codes kodiert. (Abbildung 2; Anlage 2, Kategoriensystem im Supplement S8 und S9).

### Procedure of inductive qualitative content analysis

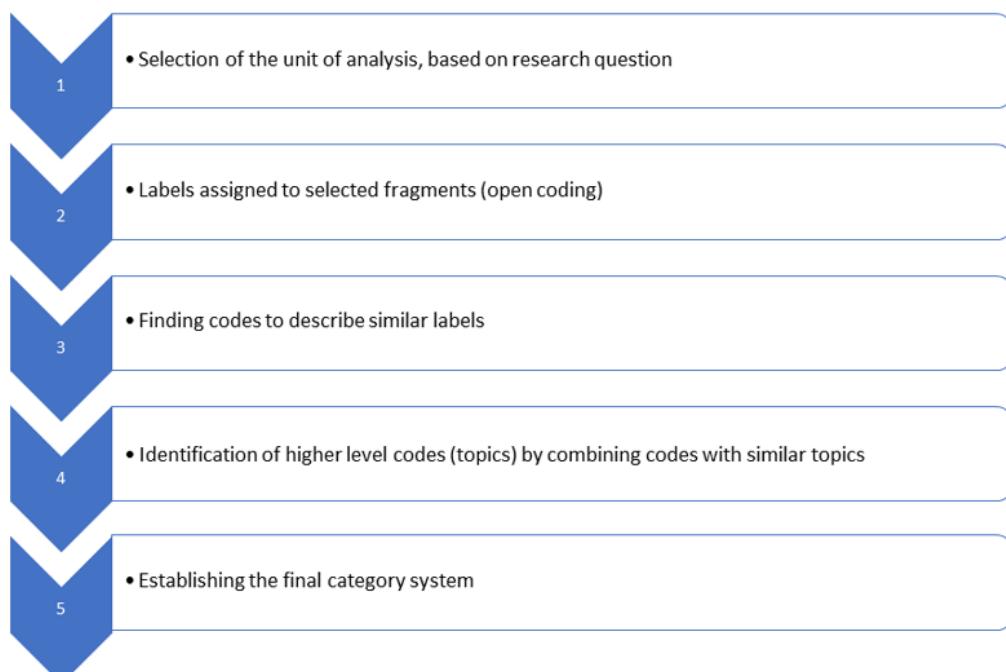


Abbildung 2: Induktive Kodierung der Interviewdaten in der DIMASOS2-Studie (Ärzt\*innen & Screeningteilnehmerinnen)

## 4 Projektergebnisse

### 4.1 Ergebnisse des quantitativen Studienteils

Die quantitative Studie wurde entsprechend dem mit dem Konsortialpartner ausgearbeiteten Studiendesign technisch durch die Konsortialführung in der Vorbereitungsphase mit der Fa. IdeeDesign geplant. Es wurde die hierfür erforderliche Technik installiert und implementiert.

Die teilnehmenden Teams wurden durch die Konsortialführung geschult und die Qualitätssicherungsmaßnahmen sowie die Systematik der Studien-Erstattungen für die teilnehmenden SE wurden ebenfalls in der Software vorbereitet.

Die an den 16 Zentren (in 15 SE) stattfindenden Untersuchungen wurden jeweils vor Ort gemäß Studienprotokoll durchgeführt. Jede teilnehmende Frau wurde lediglich einmal in die Studie entweder in IG oder KG eingebracht. (Eine Zweit-Untersuchung wurde automatisch in der Studiensoftware geblockt.)

Es waren in der IG 25.341 Frauen und in der KG 38.529 Frauen.

Die Ergebnis-Parameter wurden durch kontinuierliches Monitoring (ebenfalls Software-gestützt) regelmäßig erfasst und mit den teilnehmenden Standorten in regelmäßigen Abständen auch im Vergleich zu einem anonymisierten Benchmarking anhand der Daten aller Standorte besprochen.

Die Fragestellungen des quantitativen Studienteils sind unter der Überschrift 1.2. als Ziel 1 bis 5 zusammengefasst dargestellt.

Schlussendlich wurden Vollständigkeit und Korrektheit der Dokumentation aller teilnehmenden SE durch vorgegebene Prüf-Algorithmen seitens der Konsortialführung sowie Konsortialpartner geprüft und Fehler bei der Dokumentation, ebenso wie beim Datenexport durch enge Kommunikation der Konsortialführung mit den SE bereinigt (siehe Kapitel 3.1.7.3).

Die Ergebnisse der endgültigen statistischen Auswertung, die durch den Konsortialpartner erfolgte, sind im Folgenden unter den Punkten 4.1.1 bis 4.1.8 zusammengefasst. Die Interpretation der Ergebnisse erfolgte in enger Kooperation von Konsortialführung und Konsortialpartner.

#### **4.1.1 Voraussetzung Dichtemessung**

Die automatisierte Dichtemessung bildete eine Grundvoraussetzung für diese Studie, da durch sie ein Dichtewert ermittelt wurde, anhand dessen die Rekrutierung für IG oder KG erfolgen konnte.

Die gewählte CE-zertifizierte Software lieferte ursprünglich - wie alle kommerziellen Softwares - nur eine grobe Dichteklassifikation nach ACR-Dichte A bis D. Für die Selektion des hohen Dichtebereichs ist diese, ebenso wie eine visuelle Klassifikation schon deshalb unzureichend, da hiermit nur der obere ca. 5 %-DKD-Bereich (ACR D) oder ein Dichtebereich (ACR C+D) selektierbar wären, der den oberen ca. 50 % der Bevölkerung entspräche.

Daher wurde die Software, durch entsprechendes KI-Training angepasst, um abgestufte Dichtewerte zu ermöglichen. Mit Beginn der Studie war entsprechend Auftrag der Konsortialführung durch die Firma Densitas eine Dichteskaliierung von 0 – 100 eingerichtet worden und somit für die Studie verfügbar gemacht.

Im Rahmen der Vorbereitungsphase erfolgten ergänzende Anpassungen für noch nicht erfasste Gerätetypen (jeweils durch Training der Software an ca. 2000 MX-Fällen pro Gerätetyp).

Da es für die Dichte-Bewertung, in die sowohl Röntgendichte (Röntgenabschwächung) wie auch die Verteilung der dichten Areale im Drüsenkörper (sogenannter Masking-Faktor) eingehen, keinen verlässlichen Gold-Standard gibt, wurden zu Beginn der Rekrutierungsphase anhand von ca. 20 - 30 Fällen pro Screening-Standort die Plausibilität der Dichtekategorisierung visuell durch die teilnehmenden SE geprüft. Hierbei ergab sich mit

einzelnen Ausnahmen eine plausible und großenteils akzeptable Klassifikation, die aber nicht perfekt war. Das firmenintern durchgeführte Training der Software mit künstlicher Intelligenz (KI) berücksichtigte dabei auch Unterschiede bei MX-Geräten von 3 verschiedenen Herstellern.

Schlussendlich wurde ein Schwellenwert pro Gerätetyp ermittelt, der mindestens die oberen ca. 15 % der in der Bevölkerung vorkommenden Dichtewerte selektieren sollte. Erschwerend war, dass sich die Population naturgemäß pro Screening-Standort unterschiedlich zusammensetzte.

Noch genauere Anpassungen hätten grundlegende Untersuchungen erfordert, die nicht Thema dieser Studie waren. Im Rahmen der Studie bestätigten die regelmäßigen Stichproben dennoch größtenteils eine visuell akzeptable Selektion des oberen Dichtebereichs, weiterhin mit einzelnen Ausnahmen. Es fielen allerdings auch noch gerätetypische Abhängigkeiten der DKD-Messung bei der KI-basierten Software auf.

Die Studie wurde mit Schwellenwerten begonnen, die den o. g. DKD-Bereich selektieren sollten. Schlussendlich wurden die oberen ca. 18 % der Dichtegebiete durch die Software selektiert.

Im Rahmen der Vorbereitungsphase wurde zudem die Dichte-Software in die das Prozedere steuernde Studien-Software integriert. Diese Arbeit erfolgte durch die Fa. IdeeDesign in enger Abstimmung mit der Konsortialführung.

Die Dichtemessung an sich erwies sich als logistisch sehr gut integriert. Die verwendete Studien-Software informierte die RFA (durch Aufpoppen eines Informations-Fensters) zeitnah (meist binnen 1 - 2 Minuten) nach Abschluss der MX, wann immer der Schwellenwert überschritten war und die ScrTN damit für die Studie geeignet sein konnte.

Zeitverzögerungen durch die Dichte-Software selbst lagen im Sekundenbereich. Je nach technischer Vernetzung innerhalb der SE konnten in Einzelfällen Verzögerungen im Minutenbereich auftreten. Letztere ergaben sich jedoch nicht aus der Dichte-Software, sondern aus den Übertragungszeiten der digitalen MX-Aufnahmen ins Praxis-interne PACS-Netz, abhängig von dessen Ausstattung und Auslastung. Bei guter Vernetzung der MX-Geräte zum PACS innerhalb der SE ergaben sich keine relevanten Verzögerungen für die Information der RFA. Die Akzeptanz bei Ärzten und Personal für die gewählte Integration der Dichtemessung war gut.

Insgesamt erwies sich die software-mäßige Integration der Dichtemessung damit als praktikable implementierbare Funktion für risikoadaptiertes bevölkerungsbezogenes Screening. Mammobile waren in dieser Studie nicht eingeschlossen. Für Mammobile ohne direkte Anbindung an ein PACS-System sind abgewandelte Lösungswege noch zu prüfen.

Zusammenfassend erlaubte die gefundene Softwarelösung an den Studien-Standorten die unmittelbare Ermittlung der Dichte aus der digitalen MX. Entsprechend dem erhaltenen Dichtewert erfolgte eine Stratifizierung aller am Standort untersuchten Frauen. Der jeweils gewählte Schwellenwert erlaubte die Selektion eines DKD-Bereichs, der den oberen ca. 18 % Dichtewerten über die Studienpopulation dieser 16 Screeningstandorte entsprach.

#### **4.1.2 Deskriptive Beschreibung der Studienpopulation**

Anlage 1, Abbildung 1 zeigt ein Flussdiagramm des Einschlussverfahrens für der TN der Studie. Für die finale Analyse verbleiben 63.870 TN. Anlage 1, Tabelle 1 zeigt die Basis-Charakteristika der IG ( $n = 25.341$ ) und der KG ( $n = 38.529$ ). Das mittlere Alter der TN lag bei 55 bzw. 56 Jahren. In der IG war der Anteil der Erstuntersuchungen (31,4 % versus 26,9 %) und der Anteil mit hoher DKD ( $> 60 : 33,4$  % versus 26,8 %) etwas höher.

Insgesamt werden die Unterschiede in den Basischarakteristika zwischen den Gruppen als nur geringfügig eingeschätzt. Dennoch wurde für die weiteren Analysen ein Verfahren (IPTW) angewandt, um die strukturellen Unterschiede auszugleichen und einen verzerrungsfreien Vergleich zwischen IG und KG zu erlauben.

Deutlichere Unterschiede zeigen sich zwischen IG und KG (gewichtete Werte) in den BA- (6,58 % versus 5,36 %), BX- (3,16 % versus 1,47 %), Empfehlungen zur kurzfristigen Kontrolle (9,6/1000 versus 4,9/1000) und CA-Raten (10,68/1000 versus 7,24/1000) (Anlage 1, Tabelle 2).

#### **4.1.3 Karzinom-Entdeckungsrate für MX und MXUS, gesamt und stratifiziert nach Subgruppen (Dichte und Alter)**

Innerhalb der IG wurden 7,97 Tumore pro 1000 Frauen durch die MX und weitere 2,71 [95 %-KI 2,1; 3,4] durch die Studienmethode entdeckt (Anlage 1, Tabelle 3 und 4). Damit konnte die Rate an zusätzlichen Tumoren, wie geplant, mit einer ausreichenden Präzision festgestellt werden. Da das untere 95 %-KI für die beobachtete Rate an zusätzlichen Tumoren größer 2 ist, kann die zusätzliche Tumorentdeckungsrate durch den Ultraschall als signifikant größer als 2 pro 1000 bezeichnet werden.

In der IG wurden unter 1000 TN 10,68 Tumore der Brust diagnostiziert, in der KG waren es 7,24 / 1000 (Werte nach IPTW, Anlage 1, Tabelle 3). Damit lag die Tumorentdeckungsrate in der IG signifikant um 48 % [RD 1,48, 95 %-KI 1,25; 176] über der Tumorentdeckungsrate der KG.

Aus Anlage 1, Tabelle 3 wird weiter ersichtlich, dass der Effekt der Karzinom-Mehrentdeckungen (+48 %) insbesondere auf deutlichen Mehrentdeckungen (+67 %) bei invasiven Tumoren zurückzuführen ist, während bei *in situ* CA nur einzelne Mehrentdeckungen vorkamen (+6 %). Mit steigendem Alter nimmt der Effekt der Mehrentdeckung zu (AG 50-54: +25 % auf AG 65-69: +106 %). Bei den Folgeuntersuchungen war der Effekt der zusätzlichen Entdeckung höher als bei Erstuntersuchungen (Tumorentdeckungsrate +61 % versus +29 %). Interessanterweise nahm der Effekt der Studienmethode mit steigender DKD Gruppe etwas ab, einschränkend ist aber für den höchsten DKD-Bereich die sehr niedrige Fallzahl und die noch fehlende Zahl an Intervallkarzinomen zu nennen, die im höchsten Dichtebereich häufiger sein könnten.

#### **4.1.4 Beschreibung und Vergleich der entdeckten CA**

Insgesamt wurden in der Studie 272 CA der Brust in der IG (bei 25.341 Frauen) und 277 in der KG (mit 38.529 Frauen) diagnostiziert (Anlage 1, Tabelle 4). In der IG mit kombiniertem MXUS wurden 204 Tumore bereits durch MX und weitere 68 durch die ergänzende Studienmethode

entdeckt. Die entdeckten CA weisen insgesamt eine günstige, für das MSP typische Verteilung der Stadien auf.

Der Anteil der *in situ* CA war für die MX diagnostizierten Tumore der IG und der KG vergleichbar (27,9 % versus 30,7 %). Auch der Anteil kleiner invasiver CA (< 1 cm, entsprechend pT1a + pT1b) war mit 21,1 % versus 19,4 % gut vergleichbar.

Unter den nur durch US entdeckten Tumoren lag dieser Anteil von *in situ*-CA nur bei 5,9 %. Tendenziell sind die US-entdeckten Tumore etwas größer als die mit MX entdeckten Tumore. Dies zeigt sich auch in der Einteilung nach dem UICC Stadium.

Anlage 1, Tabelle 5 und 6 untersuchen die *in situ* und invasiven Tumore getrennt. Für die *in situ* CA (Anlage 1, Tabelle 5) sind die Ergebnisse auf Grund der geringen Fallzahl nur deskriptiv zu interpretieren. Alle vier zusätzlich nur durch die Studienmethode entdeckten *in situ* CA konnten brusterhaltend operiert werden. Zwei der 4 durch die Studienmethode mit MXUS-Interpretation entdeckten DCIS dürften allerdings auf einer unterschiedlichen Bewertung der MX durch die MX- versus MXUS-Befunder – bei diskretem Mikrokalk – beruhen.

Bei den invasiven Tumoren (Anlage 1, Tabelle 6) sind hinsichtlich Tumorausdehnung und histopathologischem Grading keine auffälligen Unterschiede zwischen IG und KG und innerhalb der IG (MX versus US) festzustellen. Einzig der Lymphknotenstatus (N-Status) ist in der IG Gruppe ungünstiger als in der KG (nodalpositiv (N+): 18,5 % versus 9,4 %) bei einem allerdings in der IG höheren Anteil von fehlenden Lymphknoten-Stadien (NX) und bezogen auf diese Untergruppen noch niedrigen Zahlen.

#### **4.1.5 Zusätzliche bildgebende und histologische Abklärungen**

In der IG wurden 6,8 % der Frauen (n = 1713), entsprechend einer Rate von 67,6 pro 1000 TN zu einer weiteren BA-untersuchung einbestellt, verglichen mit 5,3 % (n = 2024), entsprechend 52,5 / 1000, in der KG (Anlage 1, Tabelle 7). Dies stellt eine Steigerung um 28,8 % dar, die im Wesentlichen auf die zusätzliche Studienmethode zurückzuführen ist.

Auffällig ist, dass die Wiedereinbestellungsrate zur BA in der MX-Gruppe der IG von der KG-Gruppe abweicht, also mit 3,9 % deutlich niedriger ist als in der KG mit 5,3 %. Sie liegt damit um 26,4 % unter der Rate der KG. Dieses könnte auf einen Bias bei Befundung der reinen „MX“ innerhalb der IG hinweisen, der den intra-individuellen Vergleich innerhalb der IG betrifft (siehe 5.1.4 Primäre Fragestellung: Karzinom-Mehrentdeckung). Innerhalb der IG wurden von den insgesamt 6,8 % zusätzlichen BA allein durch die MX 3,9 % veranlasst und durch die Studienmethode („US“) weitere 2,9 %.

Die BX-Rate bezogen auf alle untersuchten Frauen der IG (versus der KG) ist in der IG erhöht (IG 32,4 / 1000 versus KG 14,3 / 1000) (Anlage 1, Tabelle 8), was eine mehr als doppelt so hohen BX-Rate (Faktor 2,27) oder einem Anstieg von +126,6 % entspricht. Wiederum ist die Gruppe der durch die Studienmethode entdeckten Veränderungen für den deutlichen Anstieg der BX in der IG wesentlich verantwortlich. 47,7 % der BX der IG betreffen Veränderungen, die durch die Studienmethode auffällig waren.

Erwartungsgemäß wurden mit US detektierte Veränderungen üblicherweise (wie für sonographisch sichtbare Befunde vorgeschrieben) mit Stanzbiopsien unter US Kontrolle durchgeführt. (IG-US 95,9 % versus IG-MX 54,4 % versus KG 53,7 %). Dass 14 Befunde unter

mammographischer Steuerung biopsiert wurden, ergibt sich daraus, dass auch in der IG durch nochmalige Betrachtung der MX zusätzliche rein mammographische Befunde auffallen konnten und weiter abgeklärt werden konnten.

Ähnliches wie für die BX-Empfehlungen gilt für die kurzfristigen Kontrollempfehlungen (BI-RADS 3-Befunde) (siehe Anlage 1, Tabelle 6).

Die Rate an Empfehlungen zu kurzfristigen Kontrollen liegt für die IG mit 0,97 % aller mit („MXUS“ untersuchten Frauen) ungefähr doppelt so hoch wie für die KG (alleinige „MX-Befundung) mit einer Rate von 0,5.

Anhand der in dieser Studie gewonnenen Daten ist auch (erstmalig) ein Vergleich von KG von BA-, BX- und Kontrollempfehlungen im sehr dichten Drüsengewebe (KG) versus dem Durchschnitt der gescreenten Frauen (über alle Brustdichten) im MSP möglich.

Im Vergleich zum MSP liegen diese mit durchschnittlich 5,3 % BA (versus 4,2 % durchschnittlich für Erst- und Folgerunden), mit 1,4 % BX-Empfehlungen (versus 1,1 %) deutlich höher (siehe Anlage 1, Tabelle 8). Bei Kontrollempfehlungen ergibt sich derselbe Wert wie im MSP (0,5 %). (Kooperationsgemeinschaft Mammographie 2023: Qualitätsbericht 2021).

Diese Unterschiede werden erstmalig für das MSP berichtet.

#### **4.1.6 Abschätzung der Sensitivität und Spezifität der Studienmethode, dem kombinierten Verfahren aus MX und US (MXUS)**

Sensitivität und Spezifität können derzeit nur grob abgeschätzt werden. Für eine korrekte Bestimmung dieser diagnostischen Kenngrößen müssten zunächst die Intervallkarzinome und daraus die falsch-negativen (fneg) Befunde bestimmt werden.

Im Analyseplan wurde für die Abschätzung der Sensitivität folgende Überlegung zugrunde gelegt. Bei den TN des MSP liegt die Prävalenz für Brustkrebs bei 8 / 1000. Diese Zahl ist aus den Berichten zum MSP ableitbar. Bei hoher DKD wird von einer reduzierten Sensitivität von ca. 50 % ausgegangen, es werden 4 CA pro 1000 Frauen entdeckt. Addiert man nun die durch die Studienmethode zusätzlich entdeckten CA (Annahme: 2 / 1000) würden sich insgesamt 6 / 1000 ergeben, was einer Sensitivität von 75 % entspräche (bezogen auf 8 / 1000 als "wahre" Prävalenz).

Die Abschätzung der Sensitivität nach obigen Verfahren konnte nicht auf diese Weise durchgeführt werden, da in der IG und KG die Prävalenz der entdeckten Tumore weit höher als angenommen war. In der KG wäre ein Wert von 4 / 1000 zu erwarten gewesen, beobachtet wurde ein Wert von 7,2 / 1000.

Um dennoch eine Abschätzung von Sensitivität und Spezifität durchführen zu können, wurde ein alternativer Ansatz gewählt. Hierbei wurde aus der beobachteten Prävalenz für entdeckten Brustkrebs in der KG und einer angenommenen Sensitivität der alleinigen MX die zu erwartende Gesamtprävalenz unter den TN bestimmt. Nimmt man jetzt eine Sensitivität der alleinigen MX von 50 % an und setzt diesen Wert der beobachteten Prävalenz der KG von 7,2 / 1000 gleich, ergibt sich eine Gesamtprävalenz von 14,4 / 1000 unter den TN mit hoher DKD. Da mit MXUS in der IG 10,7 Fälle pro 1000 Frauen entdeckt wurden, ergibt sich für dieses

Szenario im dichten Drüsengewebe eine Sensitivität von 74,3 % (10,7 von 14,4) und eine Spezifität von 93,1 %.

Anlage 1, Tabelle 9 bis 11 zeigen drei Szenarien für drei verschiedenen Sensitivitäten für den MX in der KG (60 %, 50 %, 40 %).

Es ergeben sich folgende Werte für die Sensitivität und Spezifität für die drei Szenarien: 89,2 % / 93,1 %, 74,3 % / 93,1 % und 59,4 % / 93,1 %. Hierbei entspricht die errechnete Sensitivität jeweils der sog. Programm-Sensitivität bezogen auf Screening + 2-Jahres-Intervall

Je nach Szenario kann die Sensitivität durch den zusätzlichen US um 20 bis 30 Prozentpunkte gesteigert werden.

#### **4.1.7 Nutzen/Schaden Verhältnis in Abhängigkeit von der Dichte**

Für die Bewertung des Nutzen-Schaden-Verhältnisses des MXUS wurde der Nutzen als zusätzlich entdeckte Tumoren und der Schaden als zusätzliche BA (Anlage 1, Abbildung 2) bzw. als zusätzliche BX (Anlage 1, Abbildung 3) operationalisiert.

Setzt man beim Vergleich von IG und KG Schaden und Nutzen ins Verhältnis, so zeigt sich folgendes:

- Die Zahl der BA-Empfehlungen pro gefundenem Karzinom der IG liegt im Vergleich zu denen der KG (6,36 : 1 versus 7,32 : 1) sogar etwas günstiger.
- Die Zahl der BX-Empfehlungen pro gefundenem Karzinom der IG liegt im Vergleich zu denen der KG (ca. 3 : 1 versus 2 : 1) höher, allerdings deutlich besser als bislang angenommen.
- Die Rate an kurzfristigen Kontrollen pro gefundenem Karzinom der IG liegt mit ca. 9:1 schlechter als in der KG mit 6,9 : 1, aber auch deutlich besser als ursprünglich erwartet.

Vergleicht man das Verhältnis von Schaden und Nutzen innerhalb der IG für MXUS versus MX, so fällt ein ungewöhnlich niedriges Verhältnis für BA-Empfehlungen in der MX-Gruppe (mit 4,9 : 1000) im Vergleich zur KG (mit 7,3 : 1000) auf. Dieses könnte auf einen Bias bei der MX-Konsensuskonferenz durch Vorwissen des US-Befunders ergeben haben.

Insgesamt ergibt sich für das Verhältnis von Schaden zu Nutzen beim intra-individuellen vergleich (MXUS versus MX) Folgendes:

- Die Zahl der BA-Empfehlungen pro gefundenem Karzinom der MXUS (6,36 : 1) liegt im Vergleich zu dem sehr niedrigen Wert für die MX-gruppe (4,9 : 1) hier ungünstiger beim Gruppenvergleich (IG versus KG, siehe oben).
- Die Zahl der BX-Empfehlungen pro gefundenem Karzinom der MXUS (3 : 1) liegt im Vergleich Wert für die MX-gruppe (2,17 : 1) höher, vergleichbar den Ergebnissen vom Kohortenvergleich (siehe oben)
- Die Zahl der Empfehlungen zu kurzfristigen Kontrollen pro gefundenem Karzinom der MXUS liegt mit ca. 9 : 1 im Vergleich zu dem Wert für die MX-gruppe (6,6 : 1) schlechter, ebenfalls vergleichbar zu den Ergebnissen des o.g. Kohortenvergleichs.

Um den Einfluss der DKD auf das Verhältnis von BA-Empfehlungen pro Karzinom und auf das Verhältnis von BX-Empfehlungen pro Karzinom im direkten Vergleich der durch die Studienmethode versus durch alleinige MX zu analysieren, wurden Ergebnisse für BA-

Empfehlungen pro Karzinom und für Biopsie-Empfehlungen pro Karzinom innerhalb der IG verglichen:

In der IG wurden (über den gesamten in der Studie untersuchten Dichtebereich gesehen) durch den zusätzlichen US etwa 29 zusätzliche BA unter 1000 Frauen verursacht, die den 2,7 zusätzlich entdeckten Tumoren pro 1000 Frauen gegenüberstehen. Erhöht man den Schwellenwert für die DKK ab der MXUS eingesetzt wird, verringert sich die Zahl der zusätzlich gefundenen CA und auch die Zahl der zusätzlichen BA. Aus Anlage 1, Abbildung 2(X-Achse: durch US zusätzlich entdeckten Tumore, Y-Achse: zusätzlichen BA) ergibt sich für die unterschiedlichen Schwellenwerte der DKK ein fast linearer Zusammenhang und damit ein relativ konstantes Nutzen-Schaden-Verhältnis. Einen "optimalen" Wert für das Nutzen-Schaden-Verhältnis lässt sich somit nicht ableiten. Die Entdeckung eines zusätzlichen Tumors ist daher (weitgehend unabhängig von der DKK) mit rund 11 (10,7) zusätzlichen BA verbunden (PPV der Abklärung durch MXUS: ca. 10 %).

Für BX ergibt sich ein ähnliches Bild (Anlage 1, Abbildung 3). Insgesamt wurden in der IG zusätzlich 15,5 BX pro 1000 Frauen durchgeführt, die wieder den 2,7 zusätzlichen CA pro 1000 gegenüberstehen. D. h., dass um einen zusätzlichen CA entdecken zu können, nur eine von 6 zusätzlichen BX zu einem positiven Befund führte. (PPV der BX-Empfehlung: ca. 17 %).

#### **4.1.8 Abhängigkeit der CA-entdeckungsrate durch MXUS je nach gesetztem Schwellenwert**

Wie im Analyseplan vorgesehen wurde die CA-entdeckungsrate für verschiedene Schwellenwerte (Anlage 1, Abbildung 2) bzw. Dichtebereiche (Anlage 1, Abbildung 4) bestimmt. Dabei zeigt sich eine relativ konstante CA-entdeckungsrate im untersuchten Dichte-Bereich, unabhängig von der DKK. Ein Trend hin zu einer höheren CA-entdeckungsrate bei höherer Dichte ist nicht erkennbar. Für den sehr kleinen Bereich mit höchsten DKK lässt sich keine verlässliche Aussage treffen, die Fallzahl ist hier zu niedrig. Verglichen mit der CA-entdeckungsrate in der Kontrollgruppe (7,2 / 1000) zeigen sich aber im untersuchten Dichtebereich fast durchgehend signifikant höhere CA-entdeckungsraten für die IG.

Betrachtet man das Alter als Einflussfaktor, so zeigt sich ab der Altersgruppe 60-64 Jahre ein deutlicher Anstieg der CA-entdeckungsrate (Anlage 2, Abbildung S4, AG: 55-59: 9,38 / 1000; AG: 65-69: 15,87 / 1000). Die CA-entdeckungsrate in der IG liegt in allen Altersgruppen über der der Kontrollgruppe. Der Abstand zwischen IG und KG, der dem Effekt von MXUS entspricht, nimmt mit zunehmendem Alter weiter zu (Anlage 2, Abbildung S6).

## **4.2 Ergebnisse des gesundheitsökonomischen Studienteils**

Mit der Rate von 2,7 pro 1.000 zusätzlich entdeckten Tumoren innerhalb der IG durch MXUS und der Annahme der Sensitivität des MX-Screenings können die Übergangswahrscheinlichkeiten des Markov-Modells (Anlage 1, Tabelle 12) für die Strategien berechnet werden. Diese sind exemplarisch für die Annahme einer Sensitivität von 60 % für das MSP (Strategie 1) in Anlage 1, Tabelle 14 dargestellt.

In Anlage 1, Tabelle 13 sind die Kosten dargestellt, die im Zusammenhang der zusätzlichen US innerhalb des MSP auftreten sowie die Kosten zur Behandlung der Tumore in Abhängigkeit

des UICC-Stadiums. Für Frauen aus der Strategie (1 = MXUS) fallen somit Kosten der zusätzlichen Studienmethode („US“) und der Befundung im Markov-Zustand „Start“ an, während für Frauen aus der Strategie (2 = MX) in diesem Schritt keine Kosten anfallen. Im Markov-Zustand „US-Screening Karzinom“ fallen für alle Frauen mit einem Karzinom, das aufgrund des US entdeckt wurde, die Kosten der BA, der BX, der Pathologie, Fallkonferenz und der Behandlung zusätzlich an. Für den Zustand „Intervallkarzinome“ unterscheiden sich die Kosten für die beiden Strategien nicht. Für Frauen aus der Strategie (1), bei denen kein Karzinom entdeckt wurde, fallen die Kosten der BA, einer möglichen BX mit Pathologie etc. in dem Zustand „Kein Karzinom oder MX- Screening-Karzinom“ an.

Unter der Strategie (1) des zusätzlichen US ergaben sich im Schnitt 163,94 € Kosten pro ScrTN. Unter der Strategie (2) ohne zusätzlichen US ergaben sich im Schnitt 118,45 € pro ScrTN. Daraus resultieren 45,59 € zusätzliche Kosten pro Person, wenn Frauen mit hoher DKD einen zusätzlichen US im Rahmen des MX-Screening erhalten würden.

Im Jahr 2021 nahmen etwa 3 Millionen Frauen am MSP teil (Kooperationsgemeinschaft Mammografie, 2023). Wenn beispielsweise diejenigen Frauen der 15 % oberen DKD-Werte einen zusätzlichen US bekommen würden (Anlage 1, Tabelle 14 Szenario 1B, entsprechend DIMASOS-Studie), entstünden zusätzliche Kosten von 20.470.500 € für die GKV.

In der deterministischen Sensitivitätsanalyse zeigte sich, dass die Häufigkeit der im US entdeckten CA mit UICC Stadium 0 den stärksten Einfluss auf die zusätzlichen Kosten durch die US-Untersuchung hatten (Anlage 1, Abbildung 5). Erhöht sich die Wahrscheinlichkeit von CA mit UICC-Stadium 0 von ursprünglich 5,9 % auf 11,5 % (95 %-Konfidenzintervall) dann sinkt die Kostendifferenz um 3 €. Die nächst größeren Einflussfaktoren auf die Kostendifferenz sind Häufigkeit von UICC I-Fällen und UICC IV-Fällen und die Entdeckungsrate der im US entdeckten CA.

Die durchgeführten Szenarioanalysen zeigen, dass eine Reduktion der Sensitivität des MX-Screenings um 10 % eine Senkung der Kostendifferenz um 0,02 € zur Folge hätte (Anlage 1, Tabelle 14). Die Analyse, in der die oberen 20 % (Szenario 1) der Frauen mit hoher DKD einen zusätzlichen US erhielten, ergäbe 27.294.000 € an zusätzlichen Kosten für die GKV (Szenario A: Sensitivität von 60 % im MSP). Dementsprechend sänken die Kosten auf 13.647.000 €, wenn nur die oberen 10 % (Szenario 3) eine zusätzliche US-Untersuchung erhalten würden und auf 6.823.500 €, wenn nur die oberen 5 % (Szenario 4) einbezogen würden.

### 4.3 Ergebnisse des qualitativen Studienteils

Es wurden insgesamt 31 Telefoninterviews geführt: 22 mit Screening-Klientinnen, 7 mit Ärzt\*innen (darunter 4 Leiter\*innen von SE) und 2 mit RFAs. Die Teilnehmenden waren in sechs unterschiedlichen SE beschäftigt: Tübingen, Hildesheim, Mainz, Bad Langensalza, Königs Wusterhausen und Alzey. Das Durchschnittsalter der ScrTN betrug 58 Jahre und das der Expert\*innen 52 Jahre. Die Interviews dauerten zwischen 20 und 90 Minuten.

#### 4.3.1 Wahrnehmung und Kenntnisse zum Thema sehr hohe DKD

Die Hälfte der Frauen war bereits vor der Teilnahme an der Studie über den Befund "dichtes Drüsengewebe" aufgeklärt. Für den Großteil dieser Frauen hatte dies bislang keine

Konsequenzen (wie z. B. zusätzliche Untersuchungen) nach sich gezogen. Die Einschätzung der befragten Expert\*innen über das allgemeine Wissen der ScrTN (Kenntnis über Befund „DKD“ sowie Effekt auf Sensitivität) variierte stark. Eine befragte Person war der Ansicht, dass kaum Kenntnisse darüber vorhanden wären. Eine weitere Person schätzte, dass es sich um ein allgemein bekanntes Thema handeln würde, die anderen Interviewten schätzten den Anteil unter den informierten Frauen auf zwischen 30 und 50 %.

#### **4.3.2 Einfluss DKD auf die diagnostische Genauigkeit des MSP**

Keiner der befragten ScrTN war bewusst, dass eine sehr hohe DKD die diagnostische Genauigkeit des MSP potentiell beeinträchtigt. Der Großteil der befragten Expert\*innen war der Ansicht, dass sehr dichtes Drüsengewebe (obere 10 %) höchst problematisch für die Begutachtung von MX sei und dieser Umstand es erforderlich mache, dass sich das Screening für betroffene Frauen im Rahmen des MSP ändern müsse, um die Sensitivität der Untersuchung zu erhöhen. Neben handgehaltenem US wurden auch die Tomosynthese oder automatisierter Brustultraschall als alternative zusätzliche Untersuchungen genannt. Als Vorteile des US wurde aufgezählt, dass US keine zusätzliche Strahlenexposition benötige und dass die Technik an sich grundsätzlich breit verfügbar sei.

#### **4.3.3 Durchführbarkeit und Praktikabilität von MXUS**

Die befragten Frauen fühlten sich durch die zusätzliche Untersuchung sicherer, aber einige Interviewpartnerinnen zweifelten nach der Studienteilnahme an, ob MX generell einen Nutzen für sie habe und ob sie eigentlich nur US bräuchten und MX eine unnötige Untersuchung für sie darstelle. Einige Frauen empfanden die Wartezeit zwischen MX und US als emotional fordernd und sie machten sich Sorgen um ihre Gesundheit.

Alle befragten Ärzt\*innen befürchteten den durch die additive Untersuchung zu erwartendem Anstieg von fpos Befunden und damit verbundene zusätzliche Abklärungs-Untersuchungen, BX sowie Stressbelastungen für Frauen mit MXUS -auffälligem Befund.

Im Rahmen der Studienlaufzeit erlebten alle befragten Expert\*innen organisatorische Grenzen (Personalmangel, Mangel an geschulten Ärzt\*innen zur Durchführung US, Zeitmangel), die einer kontinuierlichen Bereitstellung von MXUS diametral entgegenstanden und eine organisatorische Überforderung für das Screeningpersonal darstellten. Die Umsetzung von MXUS für Frauen mit hoher DKD konnte daher auch in keiner der teilnehmenden SE durchgängig realisiert werden.

Die befragten Expert\*innen äußerten zudem Bedenken dahingehend, den Frauen Informationen zur DKD mitzuteilen. Eine Ärztin gab zu bedenken, dass sich dadurch die öffentliche Wahrnehmung in Bezug auf das MSP zum negativen wandeln könnte. Andere befürchteten, dass dadurch opportunistischem (grauen) Screening Tür und Tor geöffnet werden könnte, mit zweifelhaften Ergebnissen. Denn, darin waren sich die Expert\*innen einig, die Durchführung von US erfordert Fähigkeiten und Kenntnisse, die außerhalb des organisierten Screeningbetriebs nicht selbstverständlich durch jeden beliebigen Anbieter ersetzt werden können.

Des Weiteren wurde angeführt, dass US eine hohe Konzentrationsfähigkeit erfordere und dass nach einer gewissen Anzahl von durchgeföhrten US die Qualität der Untersuchung nachlassen

könne. Viele Befragte gaben zudem an, dass zusätzlich gefundene CA ein sehr seltes Ereignis waren, was dazu führte, dass der Enthusiasmus über die Zeit abnahm und die Durchführung zusätzlicher US eine zeitliche Belastung darstellte.

## 5 Diskussion der Projektergebnisse

In diesem Bericht werden die Ergebnisse vom methodisch- epidemiologischen Aspekt und aus inhaltlich-klinischer Sicht diskutiert.

Die Studie wurde im Mixed-Methods Design mit verschiedenen Elementen (quantitativ, qualitativ, gesundheitsökonomisch) angelegt und durchgeführt. Die Diskussion folgt dieser Gliederung.

### 5.1 Quantitativer Studienanteil

#### 5.1.1 Kohorte und Studienpopulation

Beim quantitativen Teil handelte es sich um eine Kohortenstudie mit einer Kohorte von Frauen des regulären MSP mit hoher DKD, bei denen ergänzend zur qualitätsgesicherten MX ein zusätzlicher US durchgeführt wurde (IG). Die zu dieser Kohorte erhobenen Daten wurde zur Beantwortung der primären Fragestellung genutzt.

Zusätzlich war eine Kontrollkohorte geplant, zunächst aus pragmatischen Gründen als historische Kohorte, auch wenn hier ein Verzerrungsrisiko durch zeitliche Trends, ggf. unterschiedliche screenende Ärzt\*innen oder unterschiedlicher, heute nicht mehr verfügbarer MX-Geräte nicht ausgeschlossen werden konnte. Hinzu kamen unerwartete Probleme beim dafür notwendigen Bildexport aus den sehr variablen Bildarchivierungssystemen jeder teilnehmenden SE. Frühzeitig im Studienverlauf zeigte sich, dass in den teilnehmenden SE aus organisatorischen/ logistischen Gründen nur etwa 40-50 % der Frauen mit hoher DKD in die Studie eingeschlossen werden konnten. Dieses bot damit die Möglichkeit für die KG auf diejenigen TN zurückzugreifen, die aus logistischen Gründen keinen US erhalten konnten. Damit konnte statt der historischen KG eine KG gebildet werden, für die hinsichtlich des Studien-Zeitraums, der Gerätetechnik und screenender Ärzt\*innen eine wesentlich bessere Vergleichbarkeit als durch eine historische KG erwarten werden konnte.

Erfreulicherweise bestätigte sich für IG und KG eine gut vergleichbare Verteilung der Basisvariablen, wie z. B. Alter, DKD etc. Insofern sind die Vergleiche zwischen IG und KG weitgehend frei von relevanten Verzerrungen. Um die dennoch wenigen strukturellen Unterschiede zwischen den beiden Gruppen auszugleichen, wurde ein Propensity Score Wichtung-Verfahren (IPTW) zur Angleichung der beiden Kohorten angewandt.

#### 5.1.2 Auswahl des zu untersuchenden Dichtebereichs (sehr hohe DKD, obere 15 %)

Die Auswahl eines „sehr hohen“ Dichtebereichs basierte bereits bei Antragstellung auf Arbeiten von Wanders et al. (2019) und Weigel et al. (2017), die später auch ergänzt wurden durch Weigel et al. (2020). Im Vergleich zu US-amerikanischen Arbeiten betreffen beide Arbeiten streng qualitätsgesicherte europäische Screeningprogramme, für die der Wert einer zur qualitätsgesicherten Screening-MX ergänzenden US-Untersuchung zu testen war.

Andererseits beziehen sich diese Studien auf die volldigitale MX im Gegensatz zu früheren US-amerikanischen Arbeiten, die noch basierend auf digitalisierter Film-Folien-MX (mit schlechteren MX-Ergebnissen im dichten Gewebe) entstanden waren und ohnehin graues Screening betreffen.

Während die erste Arbeit von Weigel vermuten ließ, dass ein relevanter Abfall der Programmsensitivität nur den höchsten Dichtebereich (ACR D, obere 5 % der Dichtebereiche in der Bevölkerung) betrifft, musste aus der Folgearbeit (Weigel 2020) geschlossen werden, dass - ebenso wie in der Arbeit von Wanders - , der Abfall der Programmsensitivität neben dem Dichtebereich ACR D auch den oberen Bereich von ACR C betreffen dürfte. Berücksichtigt man, dass die Datenbasis der Arbeit von Weigel aus 2020 auf deutlich mehr Untersuchungen und der Beurteilung in 4 statt einer SE beruht (Weigel 2017), sollte für weitere Entscheidungen von den Arbeiten Weigel et al. (2020) und Wanders et al. (2019) ausgegangen werden. Beide bekräftigen, dass der deutlichste Abfall der Sensitivität wohl den oberen Bereich von ACR C sowie ACR D betrifft, aber insgesamt geringer ausfällt als ursprünglich für die Gruppe D angenommen, also einem Abfall um 30 %, nicht 50 % entspricht.

Für diese Studie wurde bereits bei Antragstellung gegen eine Untersuchung der gesamten Dichtegruppe ACR C+D entschieden, da dieses ca. 50 % der Gesamt- Population betroffen hätte. Ausreichend qualitätsgesicherte US-Untersuchungen (bei bekannter hoher Untersucher-Abhängigkeit) bei einem derart großen Anteil der Bevölkerung umzusetzen, erschien und erscheint weder leistbar noch realistisch und auch unter Risiko-Nutzen Aspekten medizinisch nicht sinnvoll. Daher wurde für diese Studie der obere sehr hohe Dichtebereich (obere 15 % der Dichterwerte in Population) gewählt.

Um die hierfür erforderliche genauere Dichtedifferenzierung (als mit der visuellen ACR-Klassifikation überhaupt möglich) zu erhalten, war ein automatisiertes Dichtemessprogramm erforderlich, welches abgestufte DKD-Werte (optimal 1-100) liefern konnte. Das KI-basierte Dichte-Programm der Fa. Densitas wurde selektiert, da dieses zu Studienbeginn als einziges DKD-messungen aus DICOM-Bildern (nicht aus Rohdaten) mit erheblichen logistischen Vorteilen und für MX-Geräte verschiedener Hersteller ermöglichte. Im Vergleich zu anderen Dichtemessprogrammen, die DKD basierend auf physikalischen Parametern aus den Rohdaten ermitteln, versprach das KI-basierte Messprogramm aufgrund des Trainings durch Befunder eine stärkere Mitberücksichtigung der Gewebsverteilung, die in den sog. Maskingfaktor eingeht.

In Vorversuchen wurde die Dichte-Software in Kooperation vom Hersteller derart angepasst, dass Dichtewerte in einem Bereich von 1-100 verfügbar waren. Zudem erfolgten in der Vorbereitungsphase Plausibilitätschecks der Dichterresultate. Im Rahmen der Vorbereitungsphase dieser Studie 07/2019 bis 01/2020 sollte auch ein Schwellenwert gefunden und festgelegt werden, oberhalb dessen sich eine Frau für die Studie eignete. Voruntersuchungen zeigten eine Abhängigkeit vom Gerätetyp und dessen Einstellung. Insgesamt war die optimale Wahl eines Schwellenwerts für die oberen 15 % Dichtewerte schwierig, einerseits aufgrund der Variationsbreite der in die Studie eingeschlossenen Mammographie-Geräte (3 Hersteller mit insgesamt 8 Gerätetypen – Siemens: Mammomat Inspiration, Mammomat Inspiration PRIME, Mammomat Revelation, Mammomat Fusion; Hologic: Lorad Selenia, Selenia Dimensions; GE: Senographe Essential, Senographe Pristina)

und deren Geräteeinstellungen und andererseits aufgrund der in der Population deutlich unterschiedlichen regionalen Verteilung von dichtem Drüsengewebe. Schlussendlich wurden durch die gewählten Schwellenwerte die oberen ca. 18 % Dichtewerte in der untersuchten Bevölkerung für die Studie selektiert. Eine retrospektive Eingrenzung auf einen engeren Dichtebereich auf den oberen 15 %-Dichtebereich ist prinzipiell ebenso möglich wie eine weitere Aufteilung des hohen Dichteberreichs (siehe 4.1.3 Karzinom-Entdeckungsrate für MX und MXUS, gesamt und stratifiziert nach Subgruppen (Dichte und Alter)).

### **5.1.3 Wahl der Studienmethode**

Als Studienmethode wurde explizit nicht der alleinige Ultraschall gewählt. Die Intervention in dieser Studie besteht zwar in der zum Screening ergänzenden Anwendung von Ultraschall bei mammographisch dichtem Drüsengewebe. Es wurde aber von der Studienärzten explizit gefordert, dass sie einerseits zuerst die MX unabhängig untersuchen, andererseits diese dann auch kennen, bevor sie den Ultraschall durchführen und schlussendlich die Ergebnisse beider Methoden zusammen bewerten. Einerseits wurde erwartet, dass durch Berücksichtigung der Informationen aus beiden Methoden relevante Zusatzinformationen zu gewinnen sind und damit eine bessere Diagnose möglich ist (mit positivem Effekt auf Sensitivität und Spezifität), andererseits benötigen auch die Frauen eine Gesamtwertung des Ergebnisses. Letzteres wurde damit wegen seiner klinischen Relevanz im Vergleich zur MX untersucht.

Retrospektiv gesehen scheinen die Ergebnisse dieser Studie die Wahl der Studienmethode und die durchgeführte Qualitätssicherung zu bestätigen. Obwohl ein Screening-Normal-Kollektiv untersucht wurde und obwohl bereits die qualitäts-gesicherte Screening-MX eine unerwartet hohe Detektionsrate im sehr dichten Gewebe (von 7,2/1000) ergab, lag die Rate an zusätzlich detektierten CA mit 3,45 / 1000 (siehe Anlage 1, Tabelle 3) in der IG versus der KG - auch im internationalen Vergleich - hoch (Glechner 2023). Dieses gelang mit deutlich niedrigeren Abklärungs- und BX-Raten als international durchschnittlich berichtet 2,2-fache BX-Rate in der IG statt > 3-fache BX-Rate (Glechner 2023; Melnikow 2016; Ohuchi 2016). Das Ausmaß an kurzfristigen Kontrollen ist im grauen Screening in Deutschland erfahrungsgemäß sehr hoch. Bei dieser Studie gelang es auch die Empfehlung zu kurzfristigen Kontrollen unter 1 % zu halten, obwohl sich auch diese Rate pro Tausend mit Hinzunahme der US-Untersuchung von 4,9 in der KG auf 9,6 in der IG verdoppelte (siehe Anlage 1, Tabelle 2).

### **5.1.4 Primäre Fragestellung: Karzinom-Mehrentdeckung**

Die Forschungsfragestellung der DIMASOS-Studie war, ob durch den Einsatz des zusätzlichen US bei Frauen mit sehr dichtem DKD zwei zusätzliche CA pro 1000 untersuchte Frauen entdeckt werden können. Im Studienprotokoll war festgehalten, dass die erwartete Steigerung von 2 pro 1000 mit einer Präzision von +/- 0,5 pro 1000 ermittelt werden sollte, entsprechend einem 95 %-Konfidenzintervall von 1,55 - 2,57. Hierzu wäre eine Fallzahl von 30.000 TN erforderlich. Nachdem absehbar war, dass diese Fallzahl nicht erreicht werden konnte, wurde die Fallzahlplanung angepasst und dafür die Präzision der Schätzung leicht verringert (+/- 0,55 pro 1000), was zu einer erforderlichen Fallzahl von 25.000 TN führte. Diese Fallzahl wurde erreicht.

Auf Basis der 25.341 TN in der IG ergab sich eine Brustkrebsentdeckungsrate von 10,7 pro 1000 (und damit 3,45 pro 1000 höher als in der KG), wobei innerhalb der IG

2,7 CA pro 1000 Frauen (Anlage 1, Tabelle 4) ausschließlich durch die ergänzende Studienmethode entdeckt wurden. Der beobachtete Effekt der ergänzenden Studienmethode lag somit über den erwarteten 2 / 1000. Das 95 %-KI, das von 2,1 bis 3,4 pro 1000 reicht, zeigt an, dass die Brustkrebsentdeckungsrate durch US sogar signifikant größer als 2 pro 1000 war.

Somit lässt sich die primäre Fragestellung eindeutig beantworten: Mit dem zusätzlichen US wurden bei Frauen mit hoher DKD im MSP signifikant mehr CA entdeckt. Der beobachtete Effekt lag bei 2,7 zusätzlichen CA pro 1000 Frauen [95 %-KI 2,1 - 3,2] und er lag damit statistisch signifikant über 2 / 1000.

Zusätzlich wurden Vergleiche zwischen der IG und der KG (beruhend nur auf alleiniger MX) durchgeführt. Im adjustierten Vergleich war die CA-entdeckungsrate in der IG um 3,45 pro 1000 [95 %-KI 1,90; 5,00] oder um 48 % höher. Der Effekt der Studienmethode war hier noch größer, verglichen mit der Berechnung nur auf Basis der IG.

Dass bei Vergleich von IG und KG ein noch größerer Effekt resultierte als bei intraindividuellem Vergleich innerhalb der IG, erstaunt. Wie Anlage 1, Tabelle 4 und 7 zeigen, scheint es, als ob die alleinige Befundung der MX bei reiner MX-Befundung in der IG etwas besser war als die reine MX-Befundung der KG-Gruppe. D. h., es wurden etwas mehr CA bei MX-Befundung der IG als bei der KG gefunden (8 / 1000 versus 7,2 / 1000, siehe Anlage 1, Tabelle 4), wobei bei MX-Befundung der IG auch weniger Frauen zur BA eingeladen wurden als bei völlig unabhängiger MX-Befundung in der KG (3,9 % versus 5,3 %, siehe Anlage 1, Tabelle 7).

Als Ursache hierfür sind vor allem folgende Erklärungen zu erwägen:

- Eine besonders sorgfältige Beurteilung der MX bei Befundung der IG. (Hier würde man allerdings eher erwarten, dass z. B. in der IG mehr richtig-positive und mehr fpos Befunde entstehen, was bei der fpos-Rate aber nicht der Fall war.)
- Ein gewisser Bias könnte sich trotz der beschriebenen und konsequent durchgeföhrten sequenziellen Befundung von MX, gefolgt durch MXUS bei der MX-Befundung der IG allerdings bei der Konsensuskonferenz ergeben haben; denn bei der Konsensuskonferenz war der US-Untersucher dann anwesend, wenn er - wie im Studienprotokoll zugelassen - auch als MX-Befunder agiert hatte. Möglicherweise haben dessen (US)-Vorkenntnisse oder Einschätzungen auch unbewusst die durch PVÄ und alle Befunder durchgeföhrte MX-KK beeinflusst, obwohl diese vor der MXUS-KK erfolgte.

Zusammenfassend bestätigen beide Auswertungen eine deutlich höhere Detektion nach Durchführung der Studienmethode als davor bzw. ohne die Studienmethode.

Sollte der o. g. vermutete Bias die MX-Befundung in der IG beeinflusst haben, dann sind die Ergebnisse des intra-individuellen Vergleichs beeinträchtigt, da hierdurch der Unterschied zwischen MX-Befundung und MXUS-Befundung zu Ungunsten der Studienmethode beeinflusst wird.

Ein solcher bedauerlicher Bias hat aber keinen Einfluss auf den Vergleich von IG und KG.

Da KG und IG bei sehr vergleichbaren Basisparametern völlig unabhängig voneinander befundet wurden, erachten wir den Vergleich von KG und IG als aussagekräftiger als den intra-individuellen Vergleich.

Die Sensitivität des kombinierten MXUS-Screenings lässt sich derzeit nur abschätzen, da für die Bestimmung der Programm-Sensitivität zusätzlich zu den bislang erhebbaren Daten die Intervallkarzinome aus den zwei folgenden Jahren noch fehlen. Nach Ablauf von ca. 3 Jahren nach Durchführung der Studienuntersuchung sind diese Ergebnisse zu erwarten, da bis dann Intervallkarzinome bei den Krebsregistern gemeldet sein dürften und durch Krebsregisterabgleich auch für die wichtige Folgeauswertung der Studie verfügbar werden.

In einem Szenario, bei dem die Sensitivität der MX ohne US bei Frauen mit sehr hoher DKD auf 50 % gesetzt wurde, ergab sich für die kombinierte Anwendung von MX und US eine Sensitivität von 74,3 %. Dies entspricht fast der im Studienprotokoll erwarteten Sensitivitätssteigerung auf 75 %.

Tatsächlich könnte vermutlich eine noch höhere Sensitivität erreichbar sein, da im hohen Dichtebereich in der aktuellsten Arbeit von Weigel (2020) für das rein mammographische MSP (beruhend auf MX alleine) bereits eine Programmsensitivität von 75 % für ACR C und von 70 % bei ACR D-Dichte berichtet wurde. Diese Ergebnisse sind deutlich besser als die bei Studien-Beantragung vorliegenden Ergebnisse für die MX-Befundung aus der ersten Studie von Weigel et al. (2017) zu erwarten waren. Ist die Ausgangssensitivität bereits im Bereich von 70 - 75 %, so wäre durch Detektion der zusätzlichen CA auch für MXUS eine noch höhere Sensitivität im MSP zu erwarten.

Die genaue Programmsensitivität und deren mögliche Steigerung durch die Studienmethode werden sich nach den genannten Krebsregisterabgleichen durch Einschluss der dann auch verfügbar gewordenen Intervallkarzinome von IG und KG an ausreichend großem Patientenkollektiv belegbar sein. (Eine entsprechende Studie zur Ergebnisauswertung der Folgejahre nach DIMASOS-Untersuchung liegt dem Innovationsfond bereits als Antrag vor.)

Von diesen Ergebnissen wird auch eine Einschätzung der biologischen Bedeutung und des prognostischen Werts der Mehrdetektion durch MXUS erwartet. Diese Informationen sind zudem von besonderer Bedeutung zur Beurteilung des Werts des ergänzenden US bei hoher DKD im Vergleich zu anderen Bildgebungsmethoden wie Tomosynthese oder MRT bei hoher DKD.

### **5.1.5 Beschreibung der zusätzlich gefundenen CA (Klassifikation, CA-Stadienverteilung)**

Wie oben berichtet, wurden in der Studie 272 CA der Brust in der IG (bei 25.341 Frauen) und 277 in der KG (mit 38.529 Frauen) diagnostiziert (Anlage 1, Tabelle 4). In der IG mit kombiniertem MXUS wurden 204 Tumore bereits durch MX und weitere 68 durch die ergänzende Studienmethode entdeckt.

Die entdeckten CA weisen insgesamt eine günstige, für das MSP typische Verteilung der Stadien auf. Die Tumorgrößen der nur mit der Studienmethode entdeckten CA erscheinen tendenziell etwas größer als die der MX-Gruppe. Der Anteil der DCIS innerhalb der nur mit der Studienmethode entdeckten Malignome lag bei nur 5,9 %.

Ersteres könnte sich erklären dadurch, dass unter den mammographisch sichtbaren Tumoren (der MX-Gruppe sowie der KG), analog zum aktuellen MSP generell wesentlich mehr Folgerundenkarzinome als Erstrundenkarzinome vorkommen (erstere sind durchschnittlich kleiner als letztere). Die nur mit der Studienmethode entdeckten CA betreffen eine US-

Prävalenzrunde, für die keine frühere US-Runde vorlag, die zudem auch in mammographischen Vorrunden nicht sichtbar waren. Es handelt sich also für den US um eine Prävalenzrunde mit mehr und größeren Tumoren als man vermutlich für US-Folgerunden erwarten würde. Welche Durchschnittgröße nur sonographisch oder MXUS-entdeckte CA bei einem US-Folgerunden-Screening hätten, lässt sich aus dieser Studie nicht ableiten, da der US nur einmalig bei dieser Studie durchgeführt wurde.

Bedenkt man, dass das Überleben nach Mammakarzinom stark von der Tumorgröße bei Entdeckung abhängt (Klinisches Münchener Tumorregister 2022), dürfte zumindest ein Teil der aktuellen Mehrdetektion aber biologisch relevanten Tumoren entsprechen. Hierdurch wäre dann auch ein relevanter Effekt für eine Reduktion von Intervallkarzinomen zu erhoffen.

Die Tatsache, dass nur 5,9 % DCIS allein mit der Studienmethode gefunden wurden, ist erwartbar, da DCIS typischerweise mammographisch durch Mikrokalk mit sehr hoher Sensitivität erkennbar ist und da generell nur wenige DCIS sonographisch sichtbar sind. Von den insgesamt 4 DCIS, die in der IG nur durch die Studienmethode (MXUS) entdeckt wurden, waren 2 DCIS (ohne Mikrokalk) nur sonographisch als echoarme Herdbefunde entdeckt. Die beiden mikrokalktragenden DCIS wurden nur vom MXUS-Befunderpaar korrekt auf der mitberücksichtigten MX erkannt und gewertet. In diesen beiden Fällen lag an der jeweils fraglichen Stelle auch retrospektiv kein sonographischer Befund vor. Die jeweils diskreten Mikroverkalkungen (Fall 1: eine kleine Mikrokalkgruppe; Fall 2: sehr feiner Mikrokalk) wurden bei der MX-Befundung durch das erste Befunderpaar vermutlich wegen ihres diskreten Erscheinungsbildes nicht erkannt bzw. fehlinterpretiert (was auf eine anscheinend auch bei hoher Qualitätssicherung unvermeidbare Varianz auch bei unabhängiger Doppelbefundung im MSP hinweist).

Wenngleich die Ergebnisse der Mehrdetektion durch die Studienmethode wegen aktuell noch fehlenden Intervallkarzinomen noch nicht endgültig interpretierbar sind, lassen die Verteilung der Tumorgrößen und des histologischen Gradings eine relevante biologische Bedeutung der mehr-entdeckten CA erwarten.

### **5.1.6 Beschreibung der falsch positiven Befunde (Wiedereinbestellungsrate, Biopsierate, Rate an kurzfristigen Kontroll-Empfehlungen (sog. BI-RADS 3-Befunde) (Erwartet war eine Biopsierate < 5 %.)**

Die Abwägung von Nutzen, hier operationalisiert als zusätzlich gefundene CA, und Schaden, hier zusätzliche BA, bzw. zusätzliche BX, ergab dass bei Vergleich von IG und KG (Anlage 1, Tabelle 7 und 8), also bei Anwendung der Studienmethode, - bezogen auf alle TN der jeweiligen Gruppe - etwas mehr BA erforderlich sind als in der KG (6,76 % versus 5,25 %). Es sind ca. 2,26 Mal so viele, also deutlich mehr BX-Empfehlungen (3,24 % in der IG) erforderlich als in der KG (1,43 %). Die Rate an Empfehlungen zu kurzfristigen Kontrollen ist in der IG (0,96 %) ca. doppelt so hoch wie in der KG (0,49 %).

Die ursprünglich vermutete BX-Rate von bis zu 5 % wurde mit 3,25 % erfreulicherweise deutlich unterschritten.

Dass die Raten an Empfehlungen zur BA in der KG (mit 5,3 % versus 4,2 %) ebenso wie die Raten an BX-Empfehlungen (mit 1,4 % versus 1,1 %) in der KG höher liegen als für die Normalpopulation im MSP erklärt sich dadurch, dass sich die KG auf den hohen Dichtebereich

beschränkt. Bei Kontrollempfehlungen ergibt sich derselbe Wert wie für die Normalpopulation im MSP (0,5 %). (Werte aus dem MSP aus: Kooperationsgemeinschaft Mammographie 2023: Qualitätsbericht 2021). Diese Unterschiede sind plausibel und wurden bereits für andere Screeningprogramme berichtet. Sie bestätigen die bekannten diagnostischen Probleme in dichtem Drüsengewebe. Dieser Erkenntnisgewinn ist für das MSP neu.

Vergleicht man nun das Verhältnis von Schaden zu Nutzen, so finden sich in der IG im Vergleich zur KG nicht nur höhere Abklärungs-, BX- und Kontroll-Empfehlungsraten, sondern auch höhere Detektionsraten (10,86 / 1000 versus 7,24 / 1000) (siehe Anlage 1, Tabelle 3).

Im Einzelnen zeigt sich Folgendes:

- Die Zahl der BA-Empfehlungen pro gefundenem Karzinom der IG liegt im Vergleich zu denen der KG (6,4 : 1 versus 7,3 : 1) sogar etwas günstiger, was sich im intra-individuellen Vergleich (siehe unten) nicht gleichermaßen ergibt. Dieses Ergebnis war an sich überraschend. Man kann sich die vergleichbaren Raten an BA-Empfehlungen pro gefundenem Karzinom dadurch erklären, dass unter Verwendung von US BA-Empfehlungen für einfache Zysten im Gegensatz zur alleinigen MX-Befundung entfallen.
- Die Zahl der BX-Empfehlungen pro gefundenem Karzinom der IG liegt im Vergleich zu denen der KG (mit ca. 3 : 1 versus 2 : 1) deutlich höher, aber auch deutlich niedriger als international publiziert (Melnikow 2016). Die Studienleitung geht davon aus, dass einerseits die verwendete Studien-Methode (systematische Befundung von US zusammen mit der vorliegenden MX und deren Voraufnahmen) wie auch die angewandte Qualitätssicherung hilfreich waren, um sowohl Sensitivität wie Spezifität zu optimieren. Dass dennoch gerade die sonographisch zusätzlich entdeckten soliden Befunde (wie Fibroadenome, komplizierte Zysten oder Papillome) öfter als bei MX-Befunden BX-Empfehlungen (oder auch kurzfristige Kontroll-Empfehlungen) verursachen, ist bekannt. Dass das Ausmaß an BX-bedürftigen Befunden doch auf ein akzeptables Ausmaß begrenzt blieb bei gleichzeitig sehr guter Detektion (+3,45 CA / 1000 TN), war sehr erfreulich, wenn auch ursprünglich nicht erwartet.

Zu erwähnen bleibt, dass die eingeforderte Qualitätssicherung mit Bildgebungsvergleich mit MX und MX-Voraufnahmen für die Ärzt\*innen noch zeitintensiver ist als die Durchführung eines alleinigen US.

- Die Rate an kurzfristigen Kontrollen pro gefundenem Karzinom der IG liegt mit ca. 9 : 1 etwas schlechter als in der KG mit 6,9 : 1. Dass diese Rate ebenfalls relativ günstig ausfiel (wenngleich etwas schlechter als für alleinige MX), war ebenfalls durch die Leitung ursprünglich nicht erwartet worden.

Vergleicht man die Ergebnisse der intraindividuellen Auswertung (MXUS versus MX) so fällt die sehr niedrige Rate an BA-Empfehlungen pro gefundenem Karzinom in der MX-Gruppe mit 4 : 1 im Vergleich zur KG (7,3 : 1) auf. BX-Empfehlungen und kurzfristige Kontrollen pro gefundenem Karzinom erwiesen sich als vergleichbar zur KG.

Für die besonders niedrige Rate an BA-Empfehlungen in der MX-Gruppe dürfte der bereit o. a. vermutete Bias bei der MX-Konsensuskonferenz durch die Vorkenntnisse des Ultraschall-Befunders verantwortlich sein (siehe 5.1.4 Primäre Fragestellung: Karzinom-

Mehrentdeckung). Er scheint sich im Wesentlichen nur auf die BA-Zahlen und auf die BA-Empfehlungen der MX-Gruppe auszuwirken, nicht auf BX-Empfehlungen oder kurzfristige Kontroll-Empfehlungen. Dort sind die Ergebnisse für die Studienmethode (MXUS) zu MX im intra-individuellen Vergleich vergleichbar zum o. a. Kohorten-Vergleich (IG versus KG).

Sieht man hiervon ab, so weisen die Ergebnisse des intra-Individuellen Vergleichs von (MXUS zu MX) und des Kohortenvergleichs (IG versus KG) in dieselbe Richtung: ungefähr Verdopplung der BX-Rate; Erhöhung der Kontrollempfehlungen bei gleicher bis erhöhter BA-Empfehlungsrate durch die Studienmethode.

**Insgesamt** sind durch die Studienmethode mehr belastende Untersuchungen (vor allem BX und kurzfristige Kontrollen) zu erwarten. Die zusätzlichen Belastungen pro gefundenem Karzinom, erscheinen aber in unserer Studie deutlich geringer als die aus der Literatur bekannten Ergebnisse auch für US erwarten ließen. So gibt Melnikow in ihrer Meta-Analyse Recall-Raten von bis zu 14% und Biopsie-Raten von bis zu 13% (Melnikow 2016) an. Glechner nennt Recall-Raten von bis 14,6% (Glechner 2023). Ohuchi, der die einzige randomisierte kontrollierte Studie mit Vergleich von MX und MXUS bei jüngeren Frauen (Alter: 40-49 Jahren) und jährlichem Screening publizierte, nennt Recall-Raten von 12,6% und BX-Raten von 4,5% (Ohuchi 2016). Bei dieser Studie sind aber, bedingt durch das jüngere Alter generell mehr Abklärungen als für Frauen unserer Studie (50-69) zu erwarten. Einschränkend ist auch darauf hinzuweisen, dass sich die Studienkollektive und die Studienqualität innerhalb der Meta-Analysen z.T. stark unterscheiden.

Die Daten unserer Studie sprechen für den hohen Wert einer strengen Qualitätssicherung und den direkten Vergleich von MX und US (auch im Verlauf), was allerdings auch den Aufwand und die Befundungszeiten deutlich erhöht.

Das ungünstigere Verhältnis von Anzahl BA bzw. Anzahl BX oder kurzfristiger Kontrollempfehlungen zu gefundenen CA war für verschiedene Dichtebereiche annähernd konstant, sodass auch keine Schwelle für die DKD angegeben werden kann, ab derer das Nutzen-Schaden-Verhältnis besonders günstig wäre.

Eine endgültige Einschätzung der Risiko-Stratifizierung durch Dichte erscheint von besonderer Bedeutung, sofern eine Individualisierung des Screeningprogramms zukünftig erwogen wird.

### **5.1.7 Ergänzende Untersuchungen zu Nutzen und Risiko, stratifiziert nach dem Dichte-Bereich (obere 5/ 10/ 15 %) MXUS**

Für die DKD ergab sich bei weiterer Unterteilung des untersuchten Dichtebereichs interessanterweise weder für die IG noch für die KG ein klarer Zusammenhang mit der CA-entdeckungsrate. Dabei nahm der Effekt der Studienmethode mit steigender DKD sogar etwas ab. Einschränkend ist aber hier die sehr niedrige Fallzahl bei sehr hoher DKD zu nennen.

Die zusätzlichen CA-befunde durch MXUS waren in allen Dichtegruppen vergleichbar und höher als in der KG.

Auch wenn die aktuelle Dichteverteilung im gesamten untersuchten Bereich keinen relevanten Einfluss auf das CA-Risiko zeigt und damit auch keine weitere Stratifizierbarkeit innerhalb des in dieser Studie untersuchten sehr hohen Dichtebereichs verspricht, könnte

dieses sich durch Ergänzung der Intervallkarzinome nach Krebsregisterabgleich noch ändern, sofern bei besonders hohen Dichtewerte mehr Intervallkarzinome auftreten sollten.

Ob hier eine generelle Begrenzung der Genauigkeit der Dichte-Stratifizierung besteht, lässt sich aktuell noch nicht feststellen.

Sofern allerdings erwogen wird, dass ergänzende Methoden bei erhöhtem Risiko oder begrenzter Sensitivität der MX eingesetzt werden, erscheint eine weitere Optimierung der Stratifizierung von sehr hoher Bedeutung. Ziel sollte sein, eine zahlenmäßig möglichst begrenzte TN-Gruppe mit höchstmöglichen Risiko zu finden. Die Begrenzung des Einsatzes zusätzlicher Methoden erscheint unverzichtbar, einerseits aus Gründen der Machbarkeit wie auch zur Reduzierung der zu zusätzlichen Nebenwirkungen, die für Ultraschall und auch MRT zu erwarten sind.

Folgende Möglichkeiten zur Verbesserung der Risikoadaptation sind zu erwägen:

- Ergänzende Optimierungen der Dichteklassifikation in Schritten von 1-100, auch in Abhängigkeit vom MX-Gerätetyp
- Optimierung der Risikostratifizierung durch Kombination verschiedener Risikofaktoren aus der MX (Dichte, Patientenalter und Hinzunahme weiterer Eigenschaften wie mammographisches Gewebepattern, Symmetrie, Mikrokalk-Verteilung). Erste Ergebnisse auf diesem Gebiet sind auch nach eigenen Erfahrungen (Eriksson 2022; Eriksson 2023; Lauritzen 2023; Vilmun 2024) vielversprechend.

Betrachtet man in dieser Studie das Alter als Einflussfaktor (siehe Anlage 1, Tabelle 3), so zeigt sich ein mit steigendem Alter zunehmender Effekt der Mehrentdeckung (AG 50-54: +25 % auf AG 65-69: +106 %). Bei den mammographischen Folgeuntersuchungen war der Effekt der zusätzlichen Entdeckung höher als bei Erstuntersuchungen (Tumorentdeckungsrate +61 % versus +29 %).

### 5.1.8 Klinischer Nutzen der Studienmethode

Definiert man Mehrdetektion als Nutzen, so ermöglicht der US die zusätzliche Entdeckung von Mammakarzinomen in dichtem Drüsengewebe, die allein mammographisch nicht sichtbar wären.

Aufgrund der hohen Rate an Mehrdetektionen bei hier erreichten vertretbaren Raten an Nebenwirkungen (im internationalen Vergleich sogar besonders geringer Nebenwirkungsrate) erscheint die ergänzend zur MX eingesetzte US (mit MXUS-Befundung) eine durchaus diskutable Alternative für ausgewählte Risikogruppen.

Anhand der Stadienverteilung und des Gradings ist für die zusätzliche Entdeckung invasiver CA ein vergleichbarer Nutzen wie durch die mammographische Entdeckung von invasiven Mammakarzinomen zu erhoffen. Nutzen und evtl. Schaden durch die Entdeckung von DCIS erscheinen aufgrund der wenigen durch die Studienmethode entdeckten DCIS von geringer Bedeutung

Der eigentliche biologische Wert der Mehrentdeckung lässt sich abschließend erst einschätzen, wenn durch Vergleich von IG und KG ein Effekt auf Intervallkarzinomraten

und/oder Folgerundenkarzinome sichtbar wird. Ein entsprechender Antrag für die **Überprüfung der Folgejahre nach Rekrutierung wurde beim Innovationsfonds eingereicht.**

Die endgültige Bewertung des biologischen Werts der Mehrdetektion durch Screeningmethoden ist auch als Abgrenzung zu biologisch irrelevanter Mehrentdeckung (z. B. von Überdiagnosen) von großer Bedeutung, insbesondere, wenn eine Auswahl ergänzender Methoden mit anderen Vor- und Nachteilen (Tomosynthese versus Ultraschall versus MRT) bezüglich evtl. Nebenwirkungen und auch Kosten zur Diskussion steht. (Canelo-Aybar 2021; Heywang-Köbrunner 2022; Veenhuizen 2021; Bakker 2019; Gelardi 2022)

## 5.2 Gesundheitsökonomisches Modell

Die vorliegende BIA zeigt, dass der zusätzliche US im Rahmen des MSP im Vergleich zur alleinigen MX zu durchschnittlichen Mehrkosten pro ScrTN mit MXUS in Höhe von 45,59 € führt. Die durchgeführte deterministische Sensitivitätsanalyse zeigt, dass die zusätzlichen Kosten vor allem von der Verteilung der UICC-Stadien der zusätzlich diagnostizierten Tumoren abhängen. Je mehr Tumoren mit niedrigem UICC-Stadium gefunden werden, desto geringer sind die Mehrkosten. Dies ist vor allem auf die geringeren Behandlungskosten zurückzuführen. Somit können die zusätzlichen Kosten mit einer steigenden Anzahl von Tumoren, welche mittels der zusätzlichen US-Untersuchung entdeckt werden, sinken. Ein weiterer wesentlicher Einflussfaktor ist die Anzahl der zusätzlich identifizierten Tumoren. Eine Zunahme der zusätzlichen Abklärungen oder BX hatten dabei nur einen geringen Einfluss auf die Differenz der Kosten zwischen dem MX-Screening mit und ohne US.

Unsere Ergebnisse zeigen auch, dass die Höhe der anfallenden Mehrkosten im Wesentlichen von der Anzahl der ScrTN abhängt, die den zusätzlichen US erhalten. So führen Maßnahmen, die sich auf die obersten 5 % oder 10 % der Frauen mit hoher DKD konzentrieren, zu deutlich geringeren Kosten im Vergleich zu einem breiter gefassten Einschluss von Frauen.

Ein wesentlicher Unsicherheitsfaktor in der durchgeführten Analyse bestand in der unzureichenden Kenntnis über die Anzahl und die Verteilung der Stadien der zu erwartenden Intervallkarzinome. Die Annahmen darüber basieren auf internationalen Daten, sodass Abweichungen zu deutschen Raten nicht ausgeschlossen werden können.

Die BIA kann als erster Schritt zur Abschätzung der Kosten betrachtet werden. Eine umfassendere Bewertung erfordert jedoch weiterführende Analysen, beispielsweise Kosten-Nutzwert-Analysen, welche zusätzlich den Nutzen sowie die Nutzengewichtungen der Individuen berücksichtigen.

## 5.3 Qualitativer Studienteil

Für den qualitativen Teil der Studie wurden insgesamt 31 Interviews mit Studien-TN und Screeningpersonal geführt. Dabei wurden die Themen (5.2.1) Wahrnehmung und Kenntnisse zum Thema DKD, (5.2.2) Einfluss von hoher DKD auf die diagnostische Genauigkeit im MSP, und (5.2.3) Durchführbarkeit und Praktikabilität von MXUS untersucht.

### 5.3.1 Wahrnehmung und Kenntnisse zum Thema DKD

Die meisten der befragten Fachleute überschätzten das Wissen der Frauen über dichtes Brustgewebe, insbesondere das Wissen über eine damit verbundene Beeinträchtigung der Screening-Genauigkeit. Etwa die Hälfte der befragten Frauen wusste bereits vor der Studienteilnahme, dass sie dichtes Brustgewebe haben, jedoch hatte diese Information für die meisten keine nennenswerten Konsequenzen, wie etwa zusätzliche Untersuchungen. Keine der interviewten ScrTN war sich über den negativen Effekt von hoher DKD auf die diagnostische Genauigkeit im MSP bewusst. Ähnliche Ergebnisse wurden auch in einer Studie aus den USA gefunden, bei der Studien-TN über Kenntnisse zum Thema sehr hohe DKD und einem damit assoziierten höheren Brustkrebsrisiko befragt wurden. Der überwiegende Teil der Befragten war sich des Zusammenhangs nicht bewusst (Beidler 2023).

Grundsätzlich werden den TN des MSP in Deutschland keine Informationen zum Befund „dichtes Drüsengewebe“ mitgeteilt und einige befragte Expert\*innen äußerten Vorbehalte gegenüber der individuellen Information der Klientinnen bzw. der Information der Öffentlichkeit über dichtes Brustgewebe. In einer britischen Studie wurde untersucht, ob Radiologen und Gynäkologen Frauen mit hoher DKD über diesen Befund aufklären und die Autor\*innen beschrieben, dass rund 60 % keine Information über sehr hohe DKD weitergeben, da es keine validen Leitlinien zum Umgang damit gibt (Varghese 2022). Der Mangel an Leitlinien und klinischer Evidenz beeinträchtigt auch in Deutschland den Umgang von Expert\*innen mit betroffenen Frauen. Dies betrifft die Aufklärung über den Befund, das Arzt-Patientinnen-Gespräch sowie Empfehlungen für weitere diagnostische Maßnahmen. Diese Situation beeinflusst die Fähigkeit von Frauen, informierte Entscheidungen zu treffen, negativ und kann zu einem falschen Sicherheitsgefühl nach der Teilnahme am MSP führen. Die Einstufung einer Brust als „dicht“ und das Bewusstsein über die damit verbundenen möglichen Konsequenzen könnten Frauen dazu anregen, proaktiv Schritte zur Überwachung ihrer Brustgesundheit zu unternehmen. Dieses könnte die Screening-Teilnahme verbessern, stattdessen aber auch zu vermehrten IGeL-Leistungen führen und zu hohen falsch-Positiv Raten ohne gesicherte Mehrdetektion oder sogar zu geringeren Screening-Teilnahme, wegen abnehmenden Vertrauens in die Screening-MX.

### 5.3.2 Einfluss von hoher DKD auf die diagnostische Genauigkeit im MSP

Die befragten Expert\*innen schätzten, dass dichtes Brustgewebe bei etwa 10 % der ScrTN die zuverlässige Begutachtung der MX erschweren. Ähnliche Angaben zum Anteil betroffener Frauen finden sich auch in der Literatur (Boyd 2007; Weigel 2017; Gordon 2022; Advani 2021). Viele Expert\*innen vertraten daher die Meinung, dass für diese Frauen zusätzliche Untersuchungen in das MSP integriert werden sollten. Nur eine befragte Person hielt das für überflüssig und war der Ansicht, dass das bestehende Screening ausreiche und zusätzliche Maßnahmen lediglich zu mehr fpos Befunden und Überdiagnosen führen würden. Welche zusätzliche Untersuchung die geeignetste sei, um weiterführende Untersuchungen einzuleiten, darüber waren sich die Expert\*innen uneinig. Neben US (handgehalten) wurden auch Tomosynthese und automatisierter US genannt. Zwei in den USA durchgeföhrte Studien beschrieben ebenfalls Verwirrung und Kontroversen unter befragten Radiologen und anderen Ärzten hinsichtlich der wahrgenommenen Auswirkungen von hoher DKD auf die MX und der Empfehlungen für zusätzliche Untersuchungen für diese Risikogruppe (Lourenceo 2017;

Maimone 2017). Beide Studien betonten die Notwendigkeit für die Erstellung evidenzbasierter Leitlinien, um Verwirrung, Fehlbehandlungen und finanzielle Ineffizienz zu vermeiden.

### **5.3.3 Durchführbarkeit und Praktikabilität von MXUS**

Für die befragten Frauen lag die Hauptschwierigkeit für die Durchführung des US bei der Terminfindung, doch insgesamt überwog der erwartete Nutzen einer verbesserten diagnostischen Genauigkeit für die Frauen. Ein Problem besteht darin, dass der US in ländlichen Gebieten aufgrund fehlender Infrastruktur und Personalmangel nicht flächendeckend angeboten werden kann, sodass betroffene Frauen zum nächst größeren Screening-Standort reisen müssen.

Von den Expert\*innen wurde US theoretisch als geeignete Maßnahme zur Verbesserung der diagnostischen Genauigkeit bei Frauen mit hoher DKD angesehen. Allerdings berichteten alle Befragten von großen Herausforderungen bei der Implementierung aufgrund von Kapazitätsengpässen, Personalmangel und dem Bedarf an zusätzlichen Ressourcen wie Raum und zertifizierten Screenern. Die befragten Expert\*innen hoben hervor, dass nur geschulte Personen qualitätsgesicherten US anzubieten sollten, da die Zuverlässigkeit der Untersuchung von der durchführenden Person und der Qualität des US-Geräts abhänge. Zudem war der organisatorische Aufwand insgesamt hoch, da nur relativ selten CA gefunden wurden, die nicht bereits durch die MX erkannt worden waren. Auch in anderen Studien wurde hervorgehoben, dass für US geschultes Personal erforderlich ist und dass die Screening-Genauigkeit stark vom Untersucher abhängt (Lobig 2023).

### **5.3.4 Stärken und Schwächen der qualitativen Studie**

Die Studie ist eine retrospektive, qualitative Analyse der Erfahrungen von Screening-TN und Expert\*innen. Eine Schwäche der Studie sind Verzerrungen durch Recall-Bias und die sehr niedrige Zahl interviewter Personen. Zudem haben lediglich Frauen mit einem unauffälligen Befund an einem Interview teilgenommen, d. h. es gab bei den Befragten keine falsch positiven Befunde, was die Wahrnehmung der Zusatzuntersuchung bezüglich der Nutzen-Schadens-Bilanz ins Positive verzerren könnte. Eine Stärke der Studie liegt in den gewonnenen Erkenntnissen zur Durchführbarkeit von US in sechs deutschen SE sowie zu Vorwissen und Akzeptanz von US bei ScrTN. Die qualitative Studie behandelt erstmalig dieses bisher wenig erforschte Thema.

## **5.4 Zusammenfassende Bewertung der Ergebnisse**

### **5.4.1 Dichtemessung und -stratifizierung**

Die Studie zeigte, dass eine Dichte-Stratifizierung durch Datenauslesung aus der MX technisch und logistisch innerhalb des MSP im Regelfall gut möglich ist, und bei guter Vernetzung innerhalb der SE eine Stratifizierung an sich den Routinebetrieb nicht relevant stört. Zusätzliches Personal ist hierfür nicht erforderlich. Haupt-Zeitverzug dürfte im Regelfall das zusätzliche Gespräch mit der ScrTN sein, wenn eine Frau - abhängig von ihrem Dichtewert - zu ergänzenden Untersuchungen einzuladen ist.

Für den Routinebetrieb bleibt zu beachten, dass Ausfälle bei der SE-internen Vernetzung (meist nicht bedingt durch Dichtesoftware oder das Studien-Softwareprogramm an sich) allerdings auch zu Lücken bei der regelmäßigen Dichtemessung führen können. Da ein „Nacharbeiten“ mit zunehmend komplexer Technik und Involvierung verschiedener Firmen meist nicht automatisiert möglich ist und mit komplexerer Struktur zunehmend personalintensiver wird, sollte generell auch ein Ausfallkonzept bestehen.

Nicht untersucht wurde in dieser Studie eine Umsetzung für Mammobile, die teilweise nicht direkt an das Praxisnetz angeschlossen werden können (aufgrund geographisch bisweilen extrem schlechter Internet-Anbindung). Alternativen und Ausfallskonzepte sind daher auch hierfür noch erforderlich.

Erfreulicherweise erwies sich die in der Studie verwendete Technik selbst als stabil bei allerdings hervorragendem Service durch die in dieser Studie beteiligten sehr engagierten Software-Partner. Insgesamt erschien die gewählte Umsetzung als **im Routinebetrieb praktikabel**.

Die erhaltenen Dichtemessungen erwiesen sich bei **Plausibilitätschecks** in der Vorbereitungsphase als großenteils akzeptabel, aber nicht perfekt.

Für die bei dieser Studie KI-basierte Software, die basierend auf der menschlichen Befundung trainiert wird, dürfte die Forderung nach einer graduellen Abstufung (1-100) der DKD eine besondere Herausforderung dargestellt haben, zumal die visuelle Klassifikation nach ACR 1-4 oder ACR A bis D bislang nur sehr grob ist: Die Dichtegruppe D der visuellen Klassifikation umfasst z. B. die ca. 5 – 8 % höchsten DKD-Werte, während bereits die nächste Dichteklasse 40 bis 45 weitere Prozent aller Dichtemessungen beinhaltet.

Zwar liefern Dichte-Softwares, die auf Rohdaten der MX basieren, quantitative Dichtewerte, die auf physikalischen Parametern wie Röntgenabschwächung beruhen. Bislang berücksichtigen sie allerdings nicht den sogenannten Masking Faktor, der in visuell trainierte KI-Dichte-Software eingeht. Zusammenfassend gibt es für eine optimale Dichte-Stratifizierung keinen Goldstandard.

Gewünscht war die Selektion der oberen ca. 15 – 20 % höchsten DKD-Werte, die durch den Schwellenwert ungefähr gelang.

Dass mit einer Unterteilung dieses sehr hohen Dichtebereichs (in obere 5 %, obere 10 % obere 15 %) keine weitere vergleichbare graduelle Abstufung des Risikos bei den im MX oder MXUS-Screening entdeckten CA gelang, war enttäuschend. Es dürfte sich aus der oben beschriebenen Problematik erklären.

Zu untersuchen bleibt, ob bzw. wie eine Dichteklassifikation weiter verbessert werden kann (z. B. durch genauere Anpassung an den jeweiligen Typ des MX-Geräts, durch Hinzunahme physikalischer Parameter oder ähnliches). Alternativ könnte eine verbesserte Stratifizierung auch gelingen durch Kombination von weiteren Risikofaktoren (z. B. Alter, weitere in der MX enthaltene Informationen aus Radiomics oder anamnestische bzw. metabolische Daten). Erste Arbeiten zur Risikostratifizierung durch Radiomics liegen bereits vor (Eriksson 2022; Eriksson 2023; Lauritzen 2023; Vilmun 2024).

Prinzipiell könnte sich die Dichtestratifizierung im hohen Dichtebereich noch verbessern, falls die nach Krebsregisterdatenabgleich bekannten Intervallkarzinome bevorzugt in sehr dichtem Drüsengewebe auftreten.

Damit sind für die endgültige Einschätzung des Werts der Dichteklassifikation die Daten von Intervallkarzinomen erforderlich. Sie sollten für eine verbesserte Risikostratifizierung ebenso wie für die endgültige Bewertung von US im Screening (siehe 5.4.2.) – wie bereits beantragt – unbedingt ergänzt werden.

#### 5.4.2 Zusammenfassende Wertung der medizinischen Ergebnisse

Die medizinischen Ergebnisse für den ergänzenden Einsatz von „US“ wurden innerhalb dieses Projekts sowohl durch einen intra-individuellen Vergleich („MXUS“ versus „MX“ innerhalb der IG) sowie auch durch einen Kohorten-Vergleich (IG versus KG) ermittelt. Die KG ergab sich aus Untersuchungen von Frauen derselben Screeningeinheit, desselben Standorts mit selbem Befunderteam und in vergleichbarem Zeitraum wie die IG. Es handelte sich um Frauen, für deren MX eine sehr hohe DKD (wie bei der IG) innerhalb der oberen ca. 18 % höchsten Dichtewerte ermittelt wurde und denen (aus organisatorischen Gründen) kein Angebot gemacht werden konnte. Für die gewählte Methode (ergänzender US mit MXUS-Befundung) waren die Ergebnisse auch im internationalen Vergleich sehr gut.

Für den intra-individuellen Teil der Studie ergab sich ein weniger positiver Effekt als für den Kohorten-Vergleich bei der Studienmethode. So wurden beim intraindividuellen Vergleich „nur“ 2,7 / 1000 zusätzliche CA jeweils durch die Studienmethode mehr detektiert, wohingegen beim Kohortenvergleich 3,45 / 1000 CA mehr entdeckt wurden. Weiterhin ergaben sich deutlich mehr zusätzliche BA-Empfehlungen bei dem intra-individuellen Vergleich als beim Kohorten-Vergleich (+2,9 % versus +1,5 %).

Die o. g. Unterschiede ergeben sich vor allem durch eine bei der MX-Befundung der IG statistisch signifikant geringere Rate an BA-Empfehlungen (nur 3,9 % in der MX-Gruppe) im Vergleich zur KG (5,3 % BA-Empfehlungen), obwohl die MX-Befundungen in beiden Gruppen durch dieselben Teams, im selben Zeitraum und basierend auf derselben Methode (MX) im MSP erfolgten. (Die Zahl der BA-Empfehlungen sind jeweils gemittelt über mammographische Erst- und Folgerunden und betreffen – im Gegensatz zu den im MSP berechneten Raten - nur dichtes Drüsengewebe).

Der Unterschied zwischen der alleinigen MX-Befundung innerhalb der IG und der KG muss als starker Hinweis auf einen Bias bei der MX-Befundung innerhalb der IG gewertet werden, obwohl vorgeschrieben war, dass alle Erst- Zweit- und KK-Befundungen von MX immer vor den jeweiligen MXUS-Befundungen zu erfolgen haben. Ursprünglich war davon ausgegangen worden, dass die MX-Befundungen damit nicht durch die MXUS-Befundungen beeinflusst würden.

Wie unter Punkt 4.1.5. (Zusätzliche bildgebende und histologische Abklärungen) beschrieben, geht die Studienleitung inzwischen davon aus, dass trotz protokollgerechtem Vorgehen in den SE (mit sequentieller Befundung von MX gefolgt von MXUS) dieser Bias vermutlich bei der MX-KK-Befundung auftrat. Er dürfte sich auf das Vorwissen des Befunders zurückführen lassen, der zuvor den Ultraschall durchgeführt hatte, dann aber auch bei der MX-KK anwesend war. Eine vollständige Trennung der Befunderpaare für MX und MXUS (auch in der KK!) war im

Protokoll nicht gefordert, da die Anwesenheit der MX-Befunder bei der MX-KK einerseits gesetzlich vorgeschrieben ist. Andererseits hätte die Bereitstellung komplett unterschiedlicher Befunderpaare in den SEs personell auch nicht umgesetzt werden können.

Damit ist dieser vermutliche Bias beim intra-individuellen Vergleich zu Ungunsten der Studienmethode (und zugunsten der MX-Befundung innerhalb der IG) leider nicht reparabel.

Für die medizinische Einschätzung der Studienmethode ist damit der Kohorten-Vergleich als deutlich zuverlässiger anzusehen. Dieser zeichnet sich besonders durch die sehr gute Vergleichbarkeit zwischen IG und KG aus, wie auch die erhobenen Basisparameter (Anlage 1, Tabelle 1) erfreulicherweise bestätigen.

Erfreulicherweise zeigen dennoch beide Vergleiche (intra-individueller Vergleich und Kohorten-Vergleich) eine jeweils signifikante Mehrdetektion durch die Studienmethode.

Damit ist die Haupt-Fragestellung des Nachweises einer „statistisch signifikanten Mehrdetektion im MSP“ beantwortet.

Erfreulicherweise liegt die Mehrdetektion, die beim Kohorten-Vergleich erreicht wurde, im oberen Bereich (3,45/1000) der für Normalrisiko-Populationen veröffentlichten zusätzlichen Detektionsraten von 1,8/1000 bei Ohuchi, 2/1000 bei den Kohortenstudien der Metaanalyse von Glechner und 2,6-3,5/1000 bei Hussein 2023 (Ohuchi 2016, Glechner 2023, Hussein 2023).

Dass diese Mehrdetektion erreicht werden konnte bei deutlich weniger Nebenwirkungen als ursprünglich anhand der Literatur erwartet, ist äußerst erfreulich.

Betrachtet man die KG als repräsentativ für die Ergebnisse im MSP für sehr dichtes Drüsengewebe, so verursachte die Studienmethode eine Erhöhung der BX-Rate um einen Faktor von ca. 2,3, für kurzfristige Kontrollen um den Faktor von ca. 2. Die Zahl zusätzlicher Einbestellungen zur BA war nur um ca. 36 % höher als in der KG.

In der internationalen Literatur sind Raten an kurzfristigen Kontrollen, die die Frauen durchaus ebenfalls belasten, meist gar nicht erwähnt (Melnikow 2016; Ohuchi 2016, Glechner 2023). Ohuchi führte bei anscheinend guter Qualitätssicherung mehr BX pro gefundenem CA durch. Er untersuchte allerdings auch eine jüngere Population (40-49 Jahre) an Screening-TN, die bekanntermaßen mehr gutartige Veränderungen und weniger CA als höhere Altersgruppen haben.

Diese auch im internationalen Vergleich bei unserer Studie hervorragenden Ergebnisse führt die Studienleitung einerseits auf die Wahl der Studienmethode (systematischer Vergleich von US und MX mit MX-Voruntersuchungen) und andererseits auf die in der Studie durchgeführte Qualitätssicherung zurück.

Leider ist diese Methode aber mit deutlichem Mehraufwand für die Ärzt\*innen verbunden als z. B. lediglich die alleinige Befundung von US nach unauffälliger MX. Zudem werden hoch erfahrene und ausgebildete Ärzt\*innen durch eine sehr zeitintensive Methode (> 20 Minuten) gebunden, die naturgemäß mit „nur“ 3,45 entdeckten zusätzlichen Tumoren pro 1000 TN zu einem vergleichsweise geringen Nutzen führt. Bei nur 1 - 2 Minuten Arztzeit pro MX-Screeningbefundung hingegen wurden in der KG 7,2 / 1000 CA gefunden. Bei allen

diagnostischen Abklärungsuntersuchungen liegt die Effektivität des personellen Arzteinsatzes pro gefundenem CA noch weit höher (siehe auch 5.4.3. Machbarkeit).

Insgesamt bestätigt dennoch diese Studie, dass durch zusätzlichen US eine relevante Verbesserung der Detektion im dichten Drüsengewebe, auch im MSP, möglich ist. Die Nebenwirkungen sind deutlich geringer als erwartet, aber mit einer Verdopplung der BX- und Kontrollempfehlungen nicht irrelevant. Die schlussendlich erreichbare Höhe der Sensitivität lässt sich erst nach Krebsregisterabgleich und Hinzunahme der Intervallkarzinome abschätzen. In Kapitel 4.1.6. sind wir pragmatisch von einer Sensitivität von 50 % für das MSP im dichten Drüsengewebe ausgegangen und konnten anhand der Daten dieser Studie eine voraussichtliche Gesamtsensitivität von ca. 75 % abschätzen.

Sollte die Sensitivität der MX-Befundung im dichten Gewebe, wie aktuell zu vermuten, bereits über 70 % liegen (Weigel 2020), wäre nach Hinzunahme der Mehrdetektion auch eine weitere Erhöhung der Gesamtsensitivität deutlich über 75 % nach ergänzender „US“-Untersuchung zu erwarten.

Essentiell für die Beurteilung sowohl der schlussendlich erreichbaren Sensitivität wie auch der Bedeutung der Mehrdetektion bleibt die Einschätzung des **biologischen Werts** der mehr detektierten CA. Generell könnten sehr frühe und langsam wachsende CA auch zu vermehrten Überdiagnosen führen; die Mehrdetektion von nur späten Stadien könnte weiterhin auch prognostisch ungünstig bleiben.

Anhand der bislang aus dieser Studie verfügbaren Daten zeigen mehrentdeckte invasive CA eine ungefähr vergleichbare Verteilung der Tumor-Größen und Histologien wie für die CA-Entdeckung im MSP. Damit ist für die Studienmethode anhand der bisherigen Daten eine ungefähr vergleichbare biologische Wertigkeit der mehrentdeckten CA wie für die mammographisch entdeckten invasiven CA des MSP zu erhoffen. Die wenigen mit „US“ mehrentdeckten DCIS (4 von 68 Malignomen der US-IG) dürften weder bezüglich Überdiagnosen noch bezüglich Überlebens ins Gewicht fallen.

Aktuell ist die Größenverteilung der mehrentdeckten invasiven CA sogar etwas ungünstiger als die ansonsten im MSP entdeckten CA. Dieses dürfte allerdings auch dadurch erklärbar sein, dass diese Studie bezogen auf nur US-sichtbare CA eine Prävalenzrunde darstellt, während im MSP die etwas kleineren Folgerunden-(Inzidenz-)CA ca. 80 % aller entdeckten CA ausmachen (Kooperationsgemeinschaft Mammographie Evaluationsbericht 2021). Schlussendlich wäre vermutlich für „US“-Folgerunden dann wirklich eine sehr vergleichbare Verteilung wie für das MX-Screening erwartbar.

Zur Bewertung der besonders wichtigen biologischen Wertigkeit fehlen aktuell naturgemäß noch die Daten der im Folgeintervall (optimal auch in der Folgerunde) auftretenden CA. Diese werden erst verfügbar mit Ablauf des 2-Jahresintervalls nach der Untersuchung sowie nach einem weiteren Jahr für deren Erfassung im Krebsregister. Diese Information ist generell für die Beurteilung von im Screening einzusetzenden Methoden unverzichtbar. Auch für den Vergleich mit der im Screening substitutiv einsetzbaren Tomosynthese, oder anderen ergänzend einsetzbaren Verfahren wie Kontrastmittel-gestützte Verfahren ist neben der Bewertung der Machbarkeit, der methodenspezifischen Nebenwirkungen und der Kosten die Frage zur biologischen Wertigkeit essentiell.

Um die erforderliche umfassende Beurteilung des US im Screening zu ermöglichen, ist bereits eine Folgestudie zur Erfassung von Intervall- und Folgerundenkarzinomen von IG und KG geplant und beim Innovationsfonds beantragt.

#### 5.4.3 Machbarkeit

Die logistische Machbarkeit einer Dichtemessung aus der digitalen MX im laufenden MX-Screening konnte in dieser Studie sehr gut belegt werden, wenngleich die Umsetzung in Mammobilen oder Lösungsmöglichkeiten bei Technik-Ausfällen noch offenbleibt.

Die Güte der erreichbaren Dichteklassifikation bleibt zu überprüfen, zumal eine weitere Risikozunahme mit höherer Dichte anhand der bislang gefundenen Karzinomzahlen pro Dichtebereich nicht nachweisbar war. Da diese Risikoeinschätzung bislang nur die im MX oder im MXUS-Screening gefundenen CA berücksichtigt, sind noch Änderungen denkbar, falls z. B. Intervallkarzinome vermehrt in höheren Dichtebereichen vorkommen. Auch zur korrekten Einschätzung des mit steigender Dichte (oder auch anderen Parametern) verbundenen Mammakarzinom-Risikos ist der Einschluss der Intervallkarzinome in den Folgejahren nach der Screening- oder Studienuntersuchung erforderlich.

Nach bisherigen Ergebnissen scheint eine weitere Verbesserung der Risikostratifizierung wünschenswert und essentiell, um die Zahl der mit ergänzenden Methoden zu untersuchenden Frauen besser eingrenzen zu können und damit auch die Nebenwirkungen, die immer auch die hohe Zahl asymptomatischer schlussendlich brustgesunder Frauen treffen, zu reduzieren.

Eine noch weiter verbesserte Risikostratifizierung scheint zudem erforderlich, um die aktuell schwierige Machbarkeit und die aktuell niedrige Effektivität der zusätzlichen Untersuchung zu verbessern.

Möglichkeiten hierfür umfassen eine weitere Optimierung der Dichtemessung an sich und die Einbeziehung weiterer Faktoren wie Patientenalter oder anderer Drüsengewebseigenschaften, die aus der MX gewonnen werden können (Radiomics). Erste Erfolge durch derartige Risikoklassifikationen sind bereits belegt (Eriksson 2022; Eriksson 2023; Lauritzen 2023; Vilmun 2024). Die hohe Bedeutung des Patientenalters für eine Risikostratifizierung wurde auch in dieser Studie nachgewiesen (siehe 4.1.33).

Als größtes Problem erwies sich die Umsetzbarkeit ergänzender US-Untersuchungen im MSP. Wie diese Studie zeigte, setzt die Methode hohe Erfahrung bei MX und US sowie deren gemeinsame sorgfältige Analyse voraus. Sie bindet damit hoch erfahrene und ausgebildete Ärzte bei Untersuchungen mit sehr hohem Zeitaufwand (> 20 min.). Trotz der medizinisch sehr guten Detektionsrate, erweist sich wegen der langen Untersuchungszeiten bei 3,45 Mehrdetektionen pro 1000 Frauen die MXUS im Screening als deutlich weniger effektiv als z. B. die Befundung der Screening-MX. Sie ist auch weit weniger effektiv als alle Abklärungsuntersuchungen, die bei radiologischen oder klinischen Auffälligkeiten eingesetzt werden. Dieses wurde auch von den im qualitativen Studienteil befragten Ärzten so empfunden und kommuniziert.

Die Personalverfügbarkeit stellte während der Corona-Pandemie ein erhebliches Problem dar, für welches bei geplanter Altersausweitung des MSP und bei weiter bestehendem Personalmangel in den nächsten Jahren keine wesentliche Verbesserung erwartet wird.

Zu bedenken ist auch, dass die Studie von den 15 engagiertesten von 96 SE durchgeführt wurde. Da die Studie neben der IG in der Studiengruppe noch 50 % Ablehnungen sowie eine große KG hatte, wurden pro teilnehmendem Standort weniger als 5 % aller Screening-TN einem ergänzenden „US“ zugeführt. Hierbei wurde die erwartete Rekrutierung vom Großteil der 15 SE trotz hoher Motivation nicht erreicht. Weitere ca. 60 kontaktierte SE konnten nicht zur Teilnahme an der durchgehend als interessant eingestuften Studie bewegt werden. Als Hauptgründe hierfür wurden genannt Personalmangel und die als völlig unzureichend eingeschätzte Erstattung der Studienuntersuchung. Diese war an den EBM angelehnt.

Berücksichtigt man die medizinisch durchaus überzeugenden Ergebnisse der Studienmethode und die Wertschätzung durch die TN, so erscheint die Studienmethode weiterhin eine mögliche Alternative für ausgewählte TN-Gruppen mit höherem Risiko, insbesondere wenn auch eine ergänzende Kontrastmittel-gestützte Untersuchung nicht in Frage kommt.

Ziel sollte es daher sein, die Risiko-Stratifizierung im MSP weiter so zu optimieren, dass möglichst kleine TN-Gruppen mit sehr hohem Risiko (bedingt durch Dichte und ggf. weitere Faktoren) definierbar werden. Wie oben erwähnt, ist dieses in naher Zukunft durch KI-basierte Risikostratifizierung erwartbar.

Dieses Vorgehen verspricht dann nicht nur eine verbesserte Machbarkeit, sondern auch eine deutlich höhere Effektivität. Sie würde auch Nebenwirkungen für die Gesamtpopulation deutlich reduzieren können, da die Anwendung ergänzender Methoden mit erhöhten falsch-positiv-Raten auf eine kleinere Zahl an TN konzentriert werden könnte.

#### **5.4.4 Stärken und Schwächen der Gesamt-Studie:**

Diese Studie ist eine oder die erste Studie, wo Wertigkeit und Machbarkeit einer ergänzenden US-Untersuchung, voll eingegliedert in die Screening-Logistik, -qualitätssicherung und -dokumentation eines zentralisierten Screeningprogramms getestet wurde.

Eine der Stärken der Studie ist, dass dieses Real-World-Szenario in vollem Umfang im Rahmen der Studie getestet wurde und auch im Rahmen der qualitativen Studie persönliche Befragungen von Patientinnen und MitarbeiterInnen zu neuen Erkenntnissen führten.

Eine wichtige Erkenntnis ist, dass das Einschieben ergänzender Untersuchungen, vor allem des Ultraschalls zu einer erheblichen Zusatzbelastung führen und äußerst guter Planung in den SE bedürfen. Die Anwendung von automatisiertem Ultraschall im Rahmen dieser Studie wurde von keiner SE trotz Berücksichtigung im Studienprotokoll gewählt. Einerseits bestand beim nicht-ärztlichen Personal ein mindestens vergleichbarer Personalmangel wie beim ärztlichen Personal (bei gleich langer Gesamt-Untersuchungszeit von automatisiertem US für MFA plus Arzt verglichen zur ärztlichen Untersuchung mit hand gehaltenem US); andererseits waren die Studienerstattungen hierfür völlig unzureichend. Eine zukünftige Erstattung dieser Methode nach Studienende war ebenso nicht Aussicht.

Nach Einschätzung der Studienleitung, wäre ergänzender US – aufgrund des sehr hohen Personaleinsatzes von hoch erfahrenen Ärzt\*innen – bei mehr als 5 % der gescreenten Frauen

nicht umsetzbar. Mit KI-gestützter zunehmender Verbesserung der Risikoeinschätzung erscheint gemäß aktueller Literatur (Eriksson 2022; Eriksson 2023; Lauritzen 2023; Vilmun 2024) eine verbesserte Risikostratifizierung durchaus in naher Zukunft möglich. Basierend auf den medizinisch überzeugenden Ergebnissen sollte ein solches Vorgehen erwogen werden, wenn sich die Ergebnisse dieser Studie nach Erfassung der Intervall- und Folgerunden-CA bestätigen.

Ein unerwarteter Nachteil war der Stress, der für die SE zusätzlich durch die Corona-Pandemie und ihre Folgen bestand. Prinzipiell unterstreicht diese Beobachtung allerdings das Problem der personal-intensiven Ultraschalluntersuchung, zumal ein Screeningprogramm auch bezüglich unerwarteten Stresses ausreichend robust sein müsste.

Als Schwäche der Studie mag diskutiert werden, dass sie nicht als randomisiert-kontrollierte Studie geplant wurde. Allerdings wurde ein früherer entsprechender Antrag der Studienleitung bereits mit der Begründung abgelehnt, dass durch die hierfür notwendige Aufklärung und Zustimmung der KG-TN in unserem Gesundheitswesen (mit grauem Screening und zahllosen IGeL-Angeboten) ein erheblicher Bias durch Kontamination (zwischen IG und KG) erwartbar wäre. Wären Frauen der KG über den Inhalt der Studie aufgeklärt worden, hätte durchaus die Gefahr bestanden, dass ein nicht unbedeutender Anteil an Frauen der KG sich einen ergänzenden US als sog IGeL-Leistung oder im grauen Screening verschafft. Neben der Kontamination der Studiendaten hätte dieses voraussichtlich auch neben unsicherer Entdeckung zu zahllosen unnötigen Abklärungen in der KG, also relevantem Schaden, geführt.

Da die Daten der KG-TN in unserer Studie ausschließlich als anonymisierte Einzelfiles verfügbar sind und durch Auswahl der Nicht-Rekrutierungszeiten auch ein korrekter Krebsregisterabgleich möglich ist, konnte die zuständige Ethikkommission dieser Kohortenstudie erfreulicherweise zustimmen. (Ein Krebsregister-Abgleich ist im deutschen MSP für alle gescreenten Frauen gesetzlich vorgesehen.)

Durch die systematische Dichtemessung bei allen Frauen der teilnehmenden Standorte und die vollständig gleiche Behandlung und Diagnostik in der IG (abgesehen von der Studienmethode selbst) wie in der KG und im MSP an sich (gleiches Befundungs- und Abklärungsteam, Weiterbehandlung in den regional selben zertifizierten Brustzentren etc.), bestätigte sich erfreulicherweise die erwartete sehr gute Vergleichbarkeit von IG und KG. Kleinere verbleibende Unterschiede zwischen den Gruppen, wie sie am ehesten bedingt durch die freiwillige Entscheidung für oder gegen die Studien-Teilnahme noch auftreten konnten, wurden durch eine Propensity-Score-Wichtung ausgeglichen. Damit dürfte eine große Stärke dieser Studie die sehr gute Vergleichbarkeit von IG und KG darstellen.

Die nur 50 %-ige Teilnahme muss als Nachteil dieser Studie gewertet werden. Sie erklärt sich allerdings gut dadurch, dass die meisten SE für die US-Untersuchung separate Termine, meist sogar an Folgetagen anboten. Dieser Nachteil war aus Kapazitätsgründen bei den SE leider nicht vermeidbar, was sich aus dem o. g. Real-World-Szenario ergab.

Neu konnten anhand der Ergebnisse der KG erstmalig auch umfassende Informationen nicht nur zu Detektionsraten im sehr hohen Dichtebereich, sondern auch zu Abklärungsraten in diesem Dichtebereich bei reinem MX-Screening gewonnen werden. Da sich diese durchaus relevant von den durchschnittlich erreichten Werten (Durchschnitt über alle Dichtebereiche)

unterscheiden, könnte dieses auch wertvoll für eine differenzierte Qualitätsüberwachung und -bewertung der SE (im Benchmarking) sein.

Für den intra-individuellen Vergleich zwischen „MX“ und „MXUS“ muss der erkennbare Bias bezüglich der MX-Befundung als relevante Schwäche gewertet werden. Durch das parallele Studiendesign mit intra-individuellem und Kohorten-Vergleich konnte er aufgedeckt werden. Er dürfte durch das Vorwissen des MXUS-Befunders im Rahmen der Konsensus-Konferenz entstanden sein. (Letztere konnte aus personellen Gründen in den SE nicht durch komplett separate Teams erfolgen). In unserer Studie wirkte sich die deutlich bessere Bewertung der „MX“-Befundung zu Ungunsten der Studienmethode aus. Dieser Bias ist leider nicht reparabel. Zumal verschiedene andere Screening- und Diagnostik-Studien in der Literatur mit vergleichbarem Design beim intra-individuellem Vergleich veröffentlicht sind, dürfte die Erfassung und Kommunikation dieses relevanten Bias durchaus auch von wissenschaftlichem Interesse sein.

Die Bewertung der Studienmethode sollte sich damit im Wesentlichen auf den Kohortenvergleich stützen.

Eine besondere Stärke dieser Studie ist die umfassende Dokumentation und Auswertung von Detektionsraten, BA-Raten, Kontroll- und BX-Raten. Insbesondere die Nebenwirkungen sind in den meisten Studien zu ergänzendem Ultraschall nicht oder nur lückenhaft (z.B. ohne Informationen zu kurzfristigen Kontrollraten) ausgewiesen. Dass vertretbare BA-Raten und BX-Raten prinzipiell bei hoher QS mit der Studienmethode erreichbar sind, konnte durch diese Studie unseres Wissens erstmalig in einer umfangreichen westlichen Population für ein zentralisierten Screeningprogramm nachgewiesen werden.

Eine evtl. Schwäche dieser Studie ist die unsichere Bewertung der Güte einer Dichte-Klassifikation. Sie wies aber durchaus auf noch bestehende Probleme und Forschungsfragen bezüglich einer solchen Stratifizierung hin. Obwohl die berichteten Probleme hier bei KI-basierter Dichte-Klassifikation auftraten, muss darauf hingewiesen werden, dass eine visuelle Abstufung im sehr hohen Dichtebereich durch die 4 ACR-Klassen gar nicht möglich ist, mit einem im digitalen Zeitalter unvertretbaren Personaleinsatz und rechtlicher Unsicherheit für die Befunder behaftet wären. Zudem gibt es keinen Goldstandard für eine solche Dichteabstufung im hohen Dichtebereich (obere 15 - 20 %), deren Güte nur durch die schlussendlich erreichbare Risikostratifizierung überprüfbar wäre. Letzteres könnte vermutlich auch für KI-basierte Software noch optimiert werden.

## 6 Verwendung der Ergebnisse nach Ende der Förderung

Während die Machbarkeit des ergänzenden „US“ im Rahmen des MSP eine Herausforderung darstellte, konnte diese Studie eindrücklich beweisen, dass eine signifikante Erhöhung der Detektion durch ergänzenden qualitätsgesicherten MXUS im MSP bei erhöhten, aber im internationalen Vergleich noch akzeptablen Abklärungs-, Kontroll- und Biopsie-Raten möglich ist.

Um den biologischen Wert der mehr-entdeckten CA bestmöglich auch im Vergleich zu alternativen Methoden einschätzen zu können, ist die ergänzende Erfassung von

Intervallkarzinomen und Folgerundenkarzinomen unverzichtbar. Diese wird möglich nach Ablauf von ca. 3 Folge-Jahren nach MXUS-Screening.

Diese Daten werden für die Beurteilung der Eignung ergänzender Methoden im Screening, generell international gefordert. Für die deutsche TOSYMA-Studie wurden/werden diese ergänzenden Daten soeben zusammengetragen.

Diese Daten sind ebenfalls unverzichtbar für eine erste Einschätzung der tatsächlichen Effektivität und Kosteneffektivität des bei hoher DKD eingesetzten „US“.

Die Erhebung dieser Daten ist zudem Grund-Voraussetzung, um die bereits gut recherchierten Daten von IG und KG auch zur weiteren Optimierung einer Risikostratifizierung bei hoher DKD nutzen zu können. Denn, das gesammelte pseudonymisierte Bildmaterial ist zukünftig auch für die Testung und Entwicklung anderer Risikostratifizierungen nutzbar.

Aus den o. g. Gründen hat die Studienleitung eine DIMASOS-Folgestudie (DIMASAOS-FOLLOW-UP-STUDIE (DFU)) geplant, bei der die Daten zu Intervall- und Folgerundenkarzinomen aus Studien- und KG zusammengetragen werden.

Ein entsprechender Antrag (DIMASAOS-FOLLOW-UP-STUDIE (DFU), Projektantrag VSF\_TO\_2024-015, eingereicht am 22.10.2024) liegt dem Innovationsfonds bereits vor.

Im Rahmen einer soeben gestarteten Kooperation zwischen dem Referenzzentrum Mammographie München, der FFB gGmbH München und dem Institut für Medizophysik der Universität Leuven sollen weitere Verbesserungsmöglichkeiten der Dichtemessung analysiert werden.

## 7 Erfolgte bzw. geplante Veröffentlichungen

### 7.1 Geplante Veröffentlichungen

Für den qualitativen Studienteil wurde bereits eine Publikation fertiggestellt mit dem Titel: „Optimising Breast Cancer Screening in National Mammography Screening Centres: Challenges and Insights on Implementing Additional Ultrasound for Women with Dense Breast Tissue — A Qualitative Study“. Der Artikel wurde bei BMC Cancer eingereicht.

### 7.2 Folgende Veröffentlichungen sind noch geplant, bzw. in Vorbereitung:

- Design, Durchführung und medizinische Ergebnisse der DIMASOS2-Kohortenstudie (Hauptpublikation – in Vorbereitung)
- Analyse der Unterschiede zwischen einer Kohortenstudie versus einer intra-individuellen Studie am selben Patientengut und Bewertung des beobachteten Bias
- Einfluss der Brustdichte auf Detektionsrate sowie auf Abklärungs-, Kontroll- und Biopsie-Empfehlungen im MSP: Bedeutung für die Bewertung von Qualitätsparametern in verschiedenen Regionen
- Anteil, Tumorstadien und Tumorbiologie von nur durch „US“ entdeckten Malignomen bei 25.341 MXUS-Screening-Untersuchungen

- Anteil, Tumorstadien und Tumorbiologie von nur MX-sichtbaren Tumoren 25.341 MXUS-Screening-Untersuchungen
- Unterschiede bei Detektion und Abklärung (BA-, BX- und Kontrollraten) in Abhängigkeit von DKD-Dichte und Alter im hohen Dichtebereich
- Kosten-Nutzen-Analyse von „US“-Screening in dichtem Drüsengewebe
- Umsetzung und Machbarkeit einer Dichtestratifizierung im MSP
- Machbarkeit von ergänzendem US bei sehr hoher DKD im MSP
- Wert der Doppel- und Konsensus-Befundung von „US“-Untersuchungen

#### IV Literaturverzeichnis

Advani, S.M., et al., Association of Breast Density With Breast Cancer Risk Among Women Aged 65 Years or Older by Age Group and Body Mass Index. *JAMA Netw Open*, 2021. 4(8): p. e2122810.

Albert, U.S., et al., 2008 update of the guideline: early detection of breast cancer in Germany. *J Cancer Res Clin Oncol*, 2009. 135(3): p. 339-54.

Ambinder, E. B., Lee, E., Nguyen, D. L., Gong, A. J., Haken, O. J., & Visvanathan, K. (2023). Interval breast cancers versus screen detected breast cancers: a retrospective cohort study. *Academic radiology*, 30, S154-S160.

Bakker MF, de Lange SV, Pijnappel RM, Mann RM, Peeters PHM, Monninkhof EM, Emaus MJ, Loo CE, Bisschops RHC, Lobbes MBI, de Jong MDF, Duvivier KM, Veltman J, Karssemeijer N, de Koning HJ, van Diest PJ, Mali WPTM, van den Bosch MAAJ, Veldhuis WB, van Gils CH; DENSE Trial Study Group. Supplemental MRI Screening for Women with Extremely Dense Breast Tissue. *N Engl J Med*. 2019 Nov 28;381(22):2091-2102. doi: 10.1056/NEJMoa1903986. PMID: 31774954.

Beidler, L.B., et al., Perceptions of Breast Cancer Risks Among Women Receiving Mammograph Screening. *JAMA Netw Open*, 2023. 6(1): p. e2252209.

Boyd, N.F., et al., Mammographic density and the risk and detection of breast cancer. *N Engl J Med*, 2007. 356(3): p. 227-36.

Broeders, M., et al., The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies. *J Med Screen*, 2012. 19 Suppl 1: p. 14-25.

Canelo-Aybar C, Carrera L, Beltrán J, Posso M, Rigau D, Lebeau A, Gräwingholt A, Castells X, Langendam M, Pérez E, Giorgi Rossi P, Van Engen R, Parmelli E, Saz-Parkinson Z, and Alonso-Coello P. Digital Breast Tomosynthesis Compared to Diagnostic Mammographic Projections (Including Magnification) among Women Recalled at Screening Mammography: A Systematic Review for the European Commission Initiative on Breast Cancer (Ecibc). Vol. 10, vol. 7, 2021. <https://pubmed.ncbi.nlm.nih.gov/33675147/>.

Comstock CE, Gatsonis C, Newstead GM, Snyder BS, Gareen IF, Bergin JT, Rahbar H, Sung JS, Jacobs C, Harvey JA, Nicholson MH, Ward RC, Holt J, Prather A, Miller KD, Schnall MD, Kuhl CK. Comparison of Abbreviated Breast MRI vs Digital Breast Tomosynthesis for Breast Cancer Detection Among Women With Dense Breasts

Undergoing Screening. JAMA. 2020 Feb 25;323(8):746-756. doi: 10.1001/jama.2020.0572. Erratum in: JAMA. 2020 Mar 24;323(12):1194. doi: 10.1001/jama.2020.2991. PMID: 32096852; PMCID: PMC7276668.

Corsetti, V., et al., Evidence of the effect of adjunct ultrasound screening in women with mammography-negative dense breasts: interval breast cancers at 1 year follow-up. Eur J Cancer, 2011. 47(7): p. 1021-6.

Dench, E.K., et al., Confusion and Anxiety Following Breast Density Notification: Fact or Fiction? J Clin Med, 2020. 9(4).

Eriksson M, Conant EF, Kontos D, Hall P. (2022) Risk Assessment in Population-Based Breast Cancer Screening. J Clin Oncol. 2022 Jul 10;40(20):2279-2280. doi: 10.1200/JCO.21.02827. Epub 2022 Apr 22. PMID: 35452252

Eriksson M, Román M, Gräwingholt A, Castells X, Nitrosi A, Pattacini P, Heywang-Köbrunner S, Rossi PG. European validation of an image-derived AI-based short-term risk model for individualized breast cancer screening-a nested case-control study. Lancet Reg Health Eur 2023, 37:100798. DOI: 10.1016/j.lanepe.2023.100798

Gelardi F, Ragagni EM, Sollini M, Bernardi D, Chiti A.; <https://pubmed.ncbi.nlm.nih.gov/36010240/>; Diagnostics (Basel). 2022 Aug 4;12(8):1890. doi: 10.3390/diagnostics12081890. PMID: 36010240 Free PMC article. Review.

Glechner, A., et al., Mammography in combination with breast ultrasonography versus mammography for breast cancer screening in women at average risk. Cochrane Database Syst Rev, 2023. 3(3): p. Cd009632.

Gordon, P.B., Breast Density and Risk of Interval Cancers. Can Assoc Radiol J, 2022. 73(1): p. 19-20.

Heywang-Köbrunner SH, 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. 2022 Apr;32(4):2301-2312. doi: 10.1007/s00330-021-08308-8. Epub 2021 Oct 25. PMID: 34694451; PMCID: PMC8921114.

Hussein H, Abbas E, Keshavarzi S, Fazelzad R, Bukhanov K, Kulkarni S, Au F, Ghai S, Alabousi A, Freitas V. Supplemental Breast Cancer Screening in Women with Dense Breasts and Negative Mammography: A Systematic Review and Meta-Analysis. Radiology. 2023 Mar;306(3):e221785. doi: 10.1148/radiol.221785. Epub 2023 Jan 31. PMID: 36719288.

Katalinic A, Eisemann N, Kraywinkel K, Noftz MR, Hübner J. Breast cancer incidence and mortality before and after implementation of the German mammography screening program. Int J Cancer. 2020 Aug 1;147(3):709-718. doi: 10.1002/ijc.32767. Epub 2019 Dec 4. PMID: 31675126.

KBV Kassenärztliche Bundesvereinigung (2024). Einheitlicher Bewertungsmaßstab 2024.

Khan, S. A., Hernandez-Villafuerte, K., Hernandez, D., & Schlander, M. (2023). Estimation of the stage-wise costs of breast cancer in Germany using a modeling approach. Frontiers in Public Health, 10, 946544. <https://doi.org/10.3389/fpubh.2022.946544>

Klinisches Tumorregister München, Krebsdaten von 2022; [https://www.tumorregister-muenchen.de/facts/surv/sC50f\\_G-ICD-10-C50-Mammakarzinom-Frauen-Survival.pdf](https://www.tumorregister-muenchen.de/facts/surv/sC50f_G-ICD-10-C50-Mammakarzinom-Frauen-Survival.pdf); Download vom 23.01.2025

Kooperationsgemeinschaft Mammographie (2023). Jahresbericht Evaluation 2021. Deutsches Mammographie-Screening-Programm. <https://fachservice.mammo-programm.de/download/evaluationsberichte/Eval-2021-Webversion.pdf>

Kooperationsgemeinschaft Mammographie (2023). Jahresbericht Qualitätssicherung 2021. Deutsches Mammographie-Screening-Programm.

Lauby-Secretan, B., et al., Breast-cancer screening--viewpoint of the IARC Working Group. *N Engl J Med*, 2015. 372(24): p. 2353-8.

Lauritzen AD, Lillholm M, Lynge E, Nielsen M, Karssemeijer N, Vejborg I. (2024) Radiology. Early Indicators of the Impact of Using AI in Mammography Screening for Breast Cancer. 2024 Jun;311(3):e232479. doi: 10.1148/radiol.232479. PMID: 38832880

Lauritzen AD, von Euler-Chelpin MC, Lynge E, Vejborg I, Nielsen M, Karssemeijer N, Lillholm M. Assessing Breast Cancer Risk by Combining AI for Lesion Detection and Mammographic Texture. *Radiology*. 2023 Aug;308(2):e230227. doi: 10.1148/radiol.230227. PMID: 37642571.

Lobig, F., et al., Performance of Supplemental Imaging Modalities for Breast Cancer in Women With Dense Breasts: Findings From an Umbrella Review and Primary Studies Analysis. *Clin Breast Cancer*, 2023. 23(5): p. 478-490.

Loureco, A.P., R.M. DiFlorio-Alexander, and P.J. Slanetz, Breast Density Legislation in New England: A Survey Study of Practicing Radiologists. *Acad Radiol*, 2017. 24(10): p. 1265-1267.

Maimone, S., M.D. McDonough, and S.L. Hines, Breast Density Reporting Laws and Supplemental Screening- A Survey of Referring Providers' Experiences and Understanding. *Curr Probl Diagn Radiol*, 2017. 46(2): p. 105-109.

Marmot, M.G., et al., The benefits and harms of breast cancer screening: an independent review. *Br J Cancer*, 2013. 108(11): p. 2205-40.

Mayring, P., Qualitative Content Analysis. A Step-by-Step Guide. 2022, :SAGE: London, Los Angeles.

Melnikow J, Fenton JJ, Whitlock EP, Miglioretti DL, Weyrich MS, Thompson JH, and Shah K. "Supplemental Screening for Breast Cancer in Women with Dense Breasts: A Systematic Review for the U.S. Preventive Services Task Force." *Annals of internal medicine* 164, no. 4 (2016-2-16 2016): 268-78.

Niraula, S., Biswanger, N., Hu, P., Lambert, P., & Decker, K. (2020). Incidence, Characteristics, and Outcomes of Interval Breast Cancers Compared with Screening-Detected Breast Cancers. *JAMA Network Open*, 3(9), e2018179.

Ohuchi N, Suzuki A, Sobue T, Kawai M, Yamamoto S, Zheng YF, Shiono YN, Saito H, Kuriyama S, Tohno E, Endo T, Fukao A, Tsuji I, Yamaguchi T, Ohashi Y, Fukuda M, Ishida T; J-START investigator groups. Sensitivity and specificity of mammography and adjunctive ultrasonography to screen for breast cancer in the Japan Strategic Anti-cancer Randomized Trial (J-START): a randomised controlled trial. Lancet. 2016 Jan 23;387(10016):341-348. doi: 10.1016/S0140-6736(15)00774-6.

R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing. R Foundation for Statistical Computing. <https://www.R-project.org/>

Rebolj, M., et al., Addition of ultrasound to mammography in the case of dense breast tissue: systematic review and meta-analysis. Br J Cancer, 2018. 118(12): p. 1559-1570.

Sullivan, S. D., Mauskopf, J. A., Augustovski, F., Jaime Caro, J., Lee, K. M., Minchin, M., Orlewska, E., Penna, P., Rodriguez Barrios, J.-M., & Shau, W.-Y. (2014). Budget Impact Analysis—Principles of Good Practice: Report of the ISPOR 2012 Budget Impact Analysis Good Practice II Task Force. Value in Health, 17(1), 5–14.

Varghese, J., et al., Breast Density Notification: Current UK National Practice. Clin Breast Cancer, 2022. 22(1): p. e101-e107.

Veenhuizen SGA, de Lange SV, Bakker MF, Pijnappel RM, Mann RM, Monninkhof EM, Emaus MJ, de Koekkoek-Doll PK, Bisschops RHC, Lobbes MBI, de Jong MDF, Duvivier KM, Veltman J, Karssemeijer N, de Koning HJ, van Diest PJ, Mali WPTM, van den Bosch MAAJ, van Gils CH, Veldhuis WB; DENSE Trial Study Group. Supplemental Breast MRI for Women with Extremely Dense Breasts: Results of the Second Screening Round of the DENSE Trial. Radiology. 2021 May;299(2):278-286. doi: 10.1148/radiol.2021203633. Epub 2021 Mar 16. PMID: 33724062.

Vilmun BM, Napolitano G, Lauritzen A, Lynge E, Lillholm M, Nielsen MB, Vejborg I. Diagnostics (Basel). 2024 Clinical Significance of Combined Density and Deep-Learning-Based Texture Analysis for Stratifying the Risk of Short-Term and Long-Term Breast Cancer in Screening. Aug 21;14(16):1823. doi: 10.3390/diagnostics14161823. PMID: 39202310 Free PMC article.

Waldmann, A., J. Hübner, and A. Katalinic, Trends Over Time in Breast-Cancer-Specific Mortality in Germany. Dtsch Arztebl Int, 2021. 118(31-32): p. 538-539.

Wanders JO, Holland K, Veldhuis WB, Mann RM, Pijnappel RM, Peeters PH, van Gils CH, Karssemeijer N. The effect of volumetric breast density on the risk of screen-detected and interval breast cancers: a cohort study. Breast Cancer Res. 2017; 19:67-80

Weigel S, Heindel W, Dietz C, Meyer-Johann U, Graewingholt A, Hense HW. Stratifizierung des Brustkrebsrisikos hinsichtlich der Einflüsse von Alter und mammografischer Dichte. Fortschr Röntgenstr 2020; 192: 678–685

Weigel, S., et al., Digital mammography screening: sensitivity of the programme dependent on breast density. Eur Radiol, 2017. 27(7): p. 2744-2751.

Wöckel A, Festl J, Stüber T, Brust K, Stangl S, Heuschmann PU, Albert US, Budach W, Follmann M, Janni W, Kopp I, Kreienberg R, Kühn T, Langer T, Nothacker M, Scharl A, Schreer I, Link H, Engel J, Fehm T, Weis J, Welt

## DIMASOS2 (01VSF18003)

A, Steckelberg A, Feyer P, König K, Hahne A, Kreipe HH, Knoefel WT, Denkinger M, Brucker S, Lüftner D, Kubisch C, Gerlach C, Lebeau A, Siedentopf F, Petersen C, Bartsch HH, Schulz-Wendtland R, Hahn M, Hanf V, Müller-Schimpfle M, Henscher U, Roncarati R, Katalinic A, Heitmann C, Honegger C, Paradies K, Bjelic-Radisic V, Degenhardt F, Wenz F, Rick O, Hölzel D, Zaiss M, Kemper G, Budach V, Denkert C, Gerber B, Tesch H, Hirsmüller S, Sinn HP, Dunst J, Münstedt K, Bick U, Fallenberg E, Tholen R, Hung R, Baumann F, Beckmann MW, Blohmer J, Fasching PA, Lux MP, Harbeck N, Hadji P, Hauner H, Heywang-Köbrunner S, Huober J, Hübner J, Jackisch C, Loibl S, Lück HJ, von Minckwitz G, Möbus V, Müller V, Nöthlings U, Schmidt M, Schmutzler R, Schneeweiss A, Schütz F, Stickeler E, Thomssen C, Untch M, Wesselmann S, Bücker A, Krockenberger M. Interdisciplinary Screening, Diagnosis, Therapy and Follow-up of Breast Cancer. Guideline of the DGGG and the DKG (S3-Level, AWMF Registry Number 032/045OL, December 2017) - Part 1 with Recommendations for the Screening, Diagnosis and Therapy of Breast Cancer. Geburtshilfe Frauenheilkd. 2018 Oct;78(10):927-948. doi: 10.1055/a-0646-4522. Epub 2018 Oct 19. PMID: 30369626; PMCID: PMC6202580.

Wu, T. and L.J. Warren, The Added Value of Supplemental Breast Ultrasound Screening for Women With Dense Breasts: A Single Center Canadian Experience. Can Assoc Radiol J. 2022. 73(1): p. 101-106.

## V Anlagen

Anlage 1: Abbildungen und Tabellen (Sperrfrist bis zum 31.12.2025)

Anlage 2: Supplement Abbildungen (Sperrfrist bis zum 31.12.2025)