

# Evaluationsbericht

(gemäß Nr. 14.1 ANBest-IF)



<b>Konsortialführung:</b>	Deutsche Schmerzgesellschaft e.V.
<b>Förderkennzeichen:</b>	01NVF17049
<b>Akronym:</b>	PAIN2020
<b>Projekttitel:</b>	Schmerz: Patientenorientiert.Abgestuft.Interdisziplinär.Netzwerk
<b>Autoren:</b>	Daniel Szcotkowski, Prof. Dr. Thomas Kohlmann
<b>Förderzeitraum:</b>	01. April 2018 – 31. März 2022

## Inhaltsverzeichnis

<b>I. Abkürzungsverzeichnis.....</b>	<b>3</b>
<b>II. Abbildungsverzeichnis.....</b>	<b>3</b>
<b>III. Tabellenverzeichnis.....</b>	<b>4</b>
<b>1. Ziel der Evaluation .....</b>	<b>5</b>
1.1 Aktuelle Situation in der Versorgung.....	5
1.2 Zielstellung einer neuen Versorgungsform in der ambulanten Versorgung.....	6
1.3 Zielgruppe .....	6
<b>2. Darstellung des Evaluationsdesigns .....</b>	<b>6</b>
2.1 Evaluationsansatz 1.....	7
2.1.1 Primäre Zielgrößen.....	8
2.1.2 Sekundäre Zielgrößen .....	9
2.1.3 Ein- und Ausschlusskriterien .....	11
2.1.4 Berechnung der Stichprobengröße.....	12
2.1.5 Auswertungsvarianten .....	14
2.1.6 Qualitätssicherungsmaßnahmen .....	15
2.2 Evaluationsansatz 2.....	16
2.2.1 Sekundäre Zielgrößen .....	16
2.2.2 Ein- und Ausschlusskriterien .....	16
2.2.3 Qualitätssicherungsmaßnahmen .....	17
2.2.4 Berechnung der Stichprobengröße.....	18
2.2.5 Auswertungsvarianten .....	18
2.2.6 Matching einer weiteren Kontrollgruppe aus dem WDWH der BARMER.....	19
<b>3. Ergebnisse der Evaluation .....</b>	<b>23</b>
3.1 Evaluationsansatz 1.....	23
3.1.1 Fallzahlen .....	23
3.1.2 Drop-Outs.....	25
3.1.3 Stichprobenmerkmale.....	26
3.1.4 Veränderung der primären Zielgrößen im Zeitverlauf .....	29
3.1.5 Ergänzende Analysen .....	32
3.1.6 Veränderung der sekundären Zielgrößen im Zeitverlauf .....	38
3.2 Evaluationsansatz 2.....	41
3.2.1 Kosten für Interventions- und Kontrollgruppe .....	41
3.2.2 Kosten der gematchten Stichprobe aus dem WDWH der BARMER.....	44

Akronym: PAIN2020

Förderkennzeichen: 01NVF17049

<b>4. Schlussfolgerungen und Empfehlungen des Evaluators .....</b>	<b>51</b>
<b>5. Literaturverzeichnis .....</b>	<b>52</b>
<b>6. Anhang .....</b>	<b>54</b>
<b>7. Anlagen .....</b>	<b>54</b>

## I. Abkürzungsverzeichnis

<b>AU</b>	Arbeitsunfähigkeit
<b>B</b>	Regressionskoeffizient
<b>CI</b>	Konfidenzintervall
<b>DASS</b>	Depression-Angst-Stress-Skala
<b>IG</b>	Interventionsgruppe
<b>IMA</b>	Interdisziplinäres multimodales Assessment
<b>IMST</b>	Interdisziplinäre multimodale Schmerztherapie
<b>KG</b>	Kontrollgruppe
<b>MFHW</b>	Marburger Fragebogen für das habituelle Wohlbefinden
<b>MIN</b>	Minimum
<b>MAX</b>	Maximum
<b>MW</b>	Mittelwert
<b>NRS</b>	Numerische Ratingskala
<b>OR</b>	Odds Ratio
<b>SBF</b>	Schmerzbedingte Funktionseinschränkung
<b>SBL</b>	Schmerzbeschreibungsliste
<b>SE</b>	Standardfehler
<b>SD</b>	Standardabweichung
<b>SI</b>	Schmerzintensität
<b>VR-12 (MCS)</b>	Veterans RAND 12-Item Health Survey, Mental Component Summary
<b>VR-12 (PCS)</b>	Veterans RAND 12-Item Health Survey, Physical Component Summary
<b>ZUF-8</b>	Fragebogen zur Patientenzufriedenheit

## II. Abbildungsverzeichnis

Abb. 1: Evaluationskonzept.....	8
Abb. 2: Herausforderungen bei der periodengerechten Zuordnung von Leistungsdaten.....	17
Abb. 3: Flussdiagramm zum Einschluss der PAIN2020-Teilnehmenden in die Evaluation (nach CONSORT) .....	24
Abb. 4: Primäre Zielgrößen (I Zufallskoeffizientenmodelle) .....	29
Abb. 5: Veränderung der primären Zielgrößen im Zeitverlauf (II Complete Cases).....	30
Abb. 6: Veränderung der Beeinträchtigungstage im Zeitverlauf (II Complete Cases) .....	31
Abb. 7: Veränderung der primären Zielgrößen im Zeitverlauf (für Studienteilnehmer mit bestehenden Schmerzen nach 6 Monaten) (II Complete Cases) .....	33
Abb. 8: Zentrumsbezogene Effektstärken bzgl. der Schmerzintensität nach 3 und 6 Monaten (III Available Cases) .....	34
Abb. 9: Zentrumsbezogene Effektstärken bzgl. der Schmerzbed. Funktionseinschränkung nach 3 und 6 Monaten.....	35
Abb. 10: Boxplots dokumentierter Datenerhebungszeitpunkte in Abhängigkeit vom Assessment-Termin.....	37
Abb. 11: Sekundäre Zielgrößen (I Zufallskoeffizientenmodelle) .....	38

Abb. 12: Histogramm der Kostenverteilung in den Bereichen ambulante Leistungen (AMB) und ambulante Krankenhausleistungen (KHAMB) ..... 49

Abb. 13: Histogramm der Kostenverteilung in den Bereichen stationäre Leistungen (KH) sowie Heil- und Hilfsmittel (HEILHILF) ..... 50

### III. Tabellenverzeichnis

Tab. 1: Technische Parameter der Simulationsrechnungen..... 13

Tab. 2: Gesamt-Fallzahlen in Zwei-Ebenen-Modellen, bei denen die Power 80 % beträgt ..... 13

Tab. 3: Rückberechnung der Brutto-Fallzahlen bei Berücksichtigung von Drop-Outs ..... 14

Tab. 4: Realisiertes Matchingmodell ..... 21

Tab. 5: Standardisierte mittlere Differenzen zwischen Interventions- und gematchter Kontrollgruppe ..... 22

Tab. 6: Vorhersage eines Studienabbruchs zwischen Aufnahme und Assessment (T1) ..... 25

Tab. 7: Vorhersage eines Drop-Outs zwischen Assessment (T1) und Nachbeobachtung nach 6 Monaten (T3)..... 25

Tab. 8: Rücklauf der Verlaufsbefragungen nach 3 (T2) und 6 Monaten (T3) je Zentrum ..... 26

Tab. 9: Stichprobenmerkmale ..... 27

Tab. 10: Veränderung der primären Zielgrößen im Zeitverlauf (II Complete Cases) ..... 30

Tab. 11: Varianzanalyse mit Messwiederholung für die primären Zielgrößen (II Complete Cases) .... 30

Tab. 12: Veränderung der primären Zielgrößen im Zeitverlauf (III Available Cases) ..... 32

Tab. 13: Anteil der Studienteilnehmer mit und ohne Schmerzen bei Nachbeobachtung ..... 32

Tab. 14: Varianzanalyse mit Messwiederholung für die primären Zielgrößen (für Studienteilnehmer mit bestehenden Schmerzen nach 6 Monaten) (II Complete Cases)..... 33

Tab. 15: Effektstärke für die primären Zielgrößen nach 6 Monaten bei Ausschluss protokollinadhärenter Fallkonstellationen (III Available Cases)..... 37

Tab. 16: Varianzanalyse mit Messwiederholung für die sekundären Zielgrößen (II Complete Cases) 40

Tab. 17: Veränderung der sekundären Zielgrößen nach 3 und 6 Monaten (III Available Cases) ..... 41

Tab. 18: Kosten 6 Monate vor und 6 Monate nach dem Assessment (Sekundärdatenanalyse mit Abrechnungsdaten der BARMER Krankenkasse) ..... 43

Tab. 19: Anteil der Studienteilnehmer mit Kosten größer 0 (Sekundärdatenanalyse) ..... 44

Tab. 20: Stichprobenmerkmale nach Matching ..... 45

Tab. 21: Schmerzbezogene Kosten (in €) in der Nachbeobachtung ..... 47

Tab. 22: Anteil an schmerzbezogenen Kosten größer 0 in der Vor- und Nachbeobachtung ..... 48

## 1. Ziel der Evaluation

### 1.1 Aktuelle Situation in der Versorgung

Die Prävalenz chronischer, nichttumorbedingter Schmerzen wird für die Bundesrepublik Deutschland auf bis zu 27% geschätzt und ist deshalb von großer gesundheitspolitischer Bedeutung (Häuser et al. 2014). 46,2 % der Versicherten der BARMER hatten im Jahr 2014 zumindest eine Diagnose mit direktem Schmerzbezug. Die häufigsten Diagnosen waren Rückenschmerzen [M54], Bauch- und Beckenschmerzen [R10], Gelenkschmerzen [M25.5], Kopfschmerzen [R10; G44] und Migräne [G43.0]. Chronische Schmerzerkrankungen mit den Diagnosen „Chronischer unbeeinflussbarer Schmerz“ [R52.1] und „Sonstiger chronischer Schmerz“ [R52.2] nahmen von 2005 bis 2014, um demografische Effekte bereinigt, um 122% auf 3.530 Betroffene je 100.000 BARMER-Versicherte zu. Dabei zeigt sich ein Anstieg mit zunehmendem Alter (Grobe et al. 2016). 2009 wurde in Deutschland der ICD-Code [F45.41] „Chronische Schmerzstörung mit somatischen und psychischen Faktoren“ eingeführt, um der bio-psycho-sozialen Bedingtheit chronischen Schmerzes Rechnung zu tragen. Diese Diagnose ist nach den Daten des BARMER Arztreports 2016 um das 2,8-fache auf 1.089 Betroffene je 100.000 Versicherte angestiegen. Mit zunehmendem Schweregrad I-IV nach Von Korff, zusammengefasst aus Schmerzstärke und Beeinträchtigung (Von Korff 2000), sinkt die gesundheitsbezogene Lebensqualität der Betroffenen und steigen die Kosten (Wenig et al. 2009). 16% der Befragten mit Rückenschmerz befanden sich im Schweregrad III oder IV, verursachten aber 62% der Kosten. Patienten mit chronischen Schmerzen zeigen häufig eine psychologisch-psychiatrische Komorbidität sowie eine erhöhte Mortalität (Grobe et al. 2016).

Zusammenfassend bedeutet dies, dass sowohl das persönliche Leiden bei Patienten mit chronifizierenden bzw. später chronischen Schmerzen erheblich zunimmt, als auch dass die Anforderungen dieser Gruppe Betroffener an das Gesundheitssystem sowie die sozialen Versorgungssysteme mit Leistungsansprüchen und Kosten steigen.

Die Therapie chronischer Schmerzen ist sowohl von einer Über-, Unter- als auch einer Fehlversorgung gekennzeichnet, deren Ursachen in der Abgrenzung von Zuständigkeitsbereichen, dem Mangel an interdisziplinären Versorgungsstrukturen sowie der unzureichenden Information, Schulung und Partizipation der Patienten beschrieben werden (Sachverständigenrat Gesundheit 2001).

**Die Über- und Fehlversorgung** betrifft sowohl Diagnostik als auch Therapie. Psychosoziale und funktionelle Risikofaktoren für eine Chronifizierung (yellow flags) werden häufig zu spät und im Verlauf noch immer zu wenig beachtet; ein einseitig somatisches Vorgehen resultiert in Überdiagnostik (insbesondere durch Bildgebung), medikamentöser Fehlversorgung (nicht indizierte Verordnung von Opioiden) und invasiven Therapiemaßnahmen (Marschall et al. 2013). Es wird über große Unterschiede zwischen der empfohlenen Behandlung (Leitlinien) und der tatsächlich angewendeten Therapie berichtet (Lang et al. 1999, Renker et al. 2009). In der Folge kommt es häufig zu ineffektiver Therapieeskalation bis hin zu Mehrfachoperationen und Frühverrentung. Operationen an der Wirbelsäule gehören zu den im „Recht auf Zweitmeinung“ beschriebenen „mengenanfälligen“ Operationen (mit 465.000 OP/p.a. unter den 20 häufigsten Operationen 2014).

In der Summe fördern die Über- und Fehlversorgung die zunehmende Fixierung des Patienten auf weitere ausschließlich unimodale, meist medizinische und somatische Behandlungsansätze. Von Seiten des Patienten entsteht ein eindimensionales Krankheitsmodell, das die Entwicklung eines angemessenen, eher biopsychosozialen Konzeptes, und damit einhergehend zielführender Umgangsformen verhindert.

**Die Unterversorgung** betrifft dagegen die multimodal geführte ambulante Versorgung mit der rechtzeitigen Berücksichtigung psychosozialer Faktoren sowohl in der Diagnostik als auch in der Therapie, einschließlich der Vernachlässigung der beruflichen Wiedereingliederung von

Personen mit chronischen Schmerzerkrankungen (Sachverständigenrat Gesundheit 2001). Mit der zu spät oder lange Zeit überhaupt nicht stattfindenden Berücksichtigung psychosozialer Faktoren gehen insbesondere die Unterversorgung mit angemessenen psychotherapeutischen oder sozialen Interventionen einher (Bundesärztekammer et al. 2017). In der alltäglichen Umsetzung ambulanter spezialisierter schmerzmedizinischer Diagnostik und Therapieansätze nach der Nationalen VersorgungsLeitlinie für diese Betroffenen zeigten sich wichtige Hürden. Diese bestehen vor allem in der rechtzeitigen Identifikation betroffener Patienten sowie in der Steuerung dieser in geeignete Therapieformen. Dabei ist es wichtig, die biopsychosoziale Wesenheit des Schmerzes und seiner Chronifizierung frühzeitig zu berücksichtigen. Eine spezialisierte schmerzmedizinische Diagnostik zu einem frühestmöglichen Zeitpunkt könnte eine solche an den Bedarfen der Patienten orientierte Steuerung gewährleisten.

## 1.2 Zielstellung einer neuen Versorgungsform in der ambulanten Versorgung

Das Hauptziel der konzipierten neuen Versorgungsform bestand in der Verbesserung der Versorgungsqualität und Versorgungseffizienz für Patienten mit chronifizierenden Schmerzen, im engeren Sinne durch:

- die Einführung eines interdisziplinären multimodalen Assessments (IMA) zur rechtzeitigen angemessenen und umfassenden Diagnostik, Information und Beratung schmerzkranker Patienten,
- eine frühzeitige leitliniengerechte Zuweisung in das Assessment über Versorger und Krankenkasse,
- eine patientenorientierte, gezielte, sektorenübergreifende Steuerung von Patienten mit Schmerzen bedarfsorientiert in eine abgestufte Versorgung unter Berücksichtigung von Chronifizierungsrisiken und durch
- eine Sicherung der Behandlungsqualität durch abgestimmte interdisziplinäre Diagnostik und Therapie inkl. deren Dokumentation.

## 1.3 Zielgruppe

Zielgruppe der konzipierten neuen Versorgungsform waren Patienten mit Schmerzen, die ein erhöhtes Risiko für eine Chronifizierung aufwiesen. Sie wurden als Population durch folgende primäre Indikatoren charakterisiert:

- Mindestalter 18 Jahre,
- akute Schmerzen seit mindestens 6 Wochen oder Schmerz-Rezidive trotz fachspezifischer Behandlung,
- Einschränkungen des Lebensvollzugs und/oder der gesundheitsbezogenen Lebensqualität durch den Schmerz.

## 2. Darstellung des Evaluationsdesigns

Während der Projektlaufzeit wurden einige Arbeitsplanänderungen am Evaluationskonzept vorgenommen. Im Einzelnen handelte es sich um die folgenden Arbeitsplanänderungen:

- Wegen einer geringeren Teilnahmebereitschaft musste die ursprünglich geplante Anzahl der teilnehmenden Zentren von 40 auf 29 reduziert werden.
- Die angenommene Drop-Out-Rate je Nachbefragungszeitpunkt wurde auf der Grundlage der bisherigen Erfahrungen von 20 % auf 30 % erhöht.

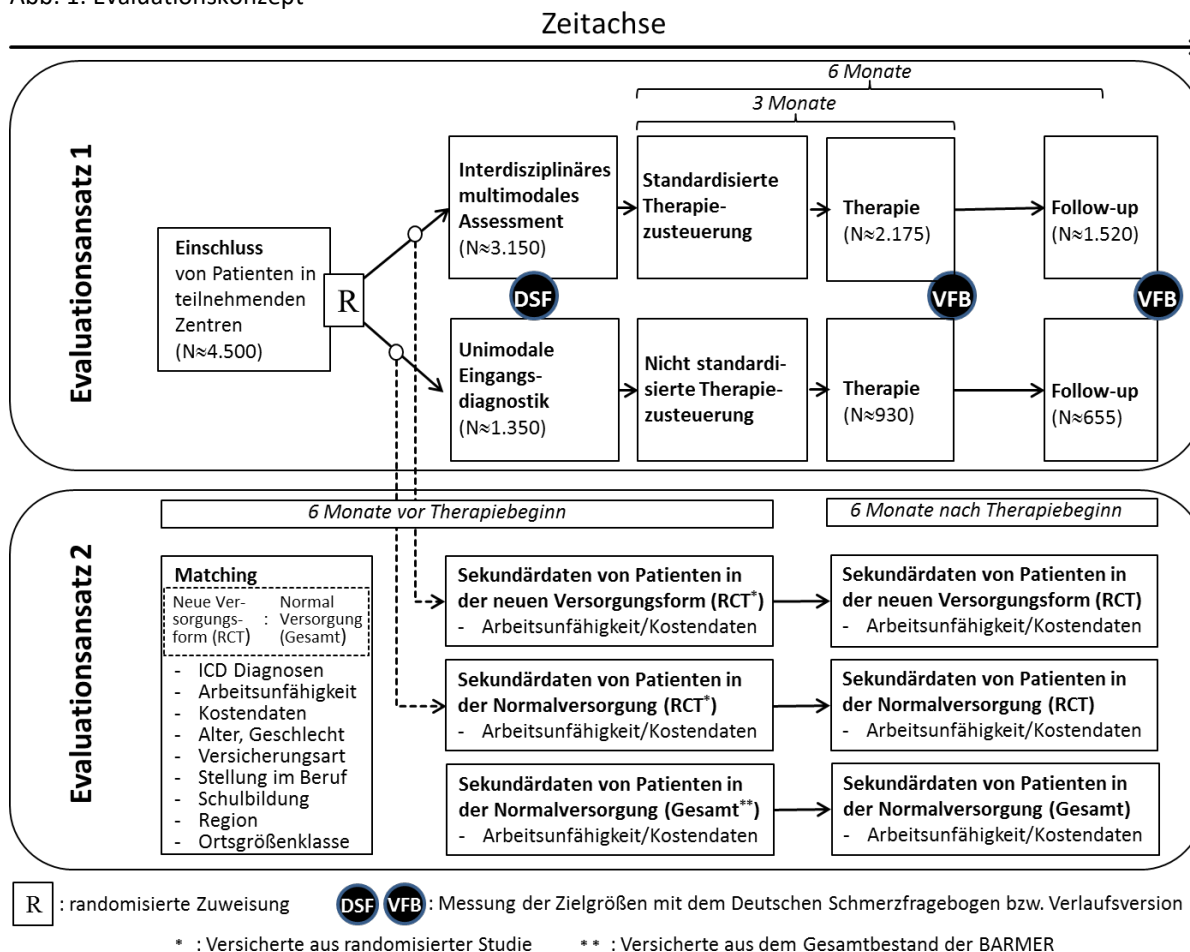
- Aufgrund dieser beiden Änderungen musste von einer insgesamt niedrigeren Fallzahl bei den rekrutierten Patienten ausgegangen werden. Die ursprünglich in der Fallzahlplanung vorgegebene Power der statistischen Tests konnte damit nicht mehr erreicht werden. Nach Beratungen im Konsortium und mit dem wissenschaftlichen Projektbeirat wurde entschieden, die Schmerzintensität und die schmerzbedingte Funktionseinschränkung weiterhin mit zwei Nachbeobachtungszeitpunkten als primäre Zielgrößen beizubehalten. Die Behandlungszufriedenheit wurde jedoch aus der Gruppe der primären Zielgrößen herausgenommen und als sekundäre Zielgröße weitergeführt. Diese Entscheidung war auch dadurch mitbegründet, dass sich bei der Erfassung der Behandlungszufriedenheit Verständnisschwierigkeiten bei den Befragten ergeben hatten.
- Die Berechnung der benötigten Brutto-Fallzahlen auf Basis dieser geänderten Annahmen ergab eine Gesamt-Brutto-Stichprobengröße von N = 4.500 Patienten. Es zeigte sich, dass mit diesem Stichprobenumfang bei den meisten der betrachteten Szenarien eine Power von 80 % erreicht werden kann.
- Für die Sekundärdatenanalyse (Evaluationsansatz 2, siehe Kapitel 2.2) ergaben die Berechnungen auf der Grundlage der reduzierten Fallzahlen, dass in diesem Bereich keine nennenswerten Auswirkungen auf die Wahrscheinlichkeit, mit der Unterschiede zwischen den Gruppen entdeckt werden können, zu erwarten sind.

## 2.1 Evaluationsansatz 1

Teilnehmende wurden nach Einschluss in **zwei Studienarme** randomisiert (**Abb. 1**), und zwar:

- zu 70% in die *Interventionsgruppe (IG)*, welche mit dem interdisziplinären multimodalen Assessment (IMA) die neue Versorgungsform und eine standardisierte Therapieeinstellung erhielt,
- zu 30% in die *Kontrollgruppe (KG)*, für die im Rahmen der schmerztherapeutischen Regelversorgung ein unimodales Eingangsdiagnostik und eine nicht standardisierte Therapieeinstellung vorgesehen war.

Abb. 1: Evaluationskonzept



Die Datenerhebung erfolgte mittels des Deutschen Schmerzfragebogens (DSF) bzw. des zugehörigen Verlaufsfragebogens (VFB). Initial wurden ausschließlich die jeweiligen Papierversionen verwendet. Ende 2020 eröffnete die Projektleitung in PAIN2020 nach Abstimmung mit dem Evaluationsinstitut den Zentren die Möglichkeit, Studienteilnehmer auch per E-Mail zu einer elektronischen Version der Verlaufsbefragung einzuladen. Hierbei wurde darauf geachtet, dass die elektronische Version in Struktur und Layout möglichst genau der Papierversion des VFB entsprach.

### 2.1.1 Primäre Zielgrößen

Im Evaluationsansatz 1 der randomisierten kontrollierten Studie wurden zwei **primäre Zielgrößen** untersucht:

- Schmerzintensität (SI) und die
- schmerzbedingte Funktionseinschränkung (SBF).

Die SI und SBF sind Bestandteile des Schweregrades nach Von Korff (2000) und werden durch jeweils drei numerische Ratingskalen erfasst. Die Endpunkte der 11-stufigen Ratingskalen sind mit „kein Schmerz“ bzw. „keine Beeinträchtigung“ (0) und „stärkster vorstellbarer Schmerz“ bzw. „völlige Beeinträchtigung“ (10) bezeichnet. Aus den Messwerten der drei zur Schmerzintensität (momentane, durchschnittliche und größte Schmerzstärke) bzw. zur Funktionseinschränkung (im Alltag, bei Freizeitaktivitäten, im Beruf und Haushalt) gehörigen Ratingskalen werden die Mittelwerte berechnet. Diese Mittelwerte bilden die beiden



primären Endpunkte SI und SBF. Die entsprechenden Fragen waren im DSF bzw. VFB enthalten.

Diese Zielgrößen wurden zu drei Messzeitpunkten erfasst, zum Zeitpunkt des Einschlusses der Teilnehmer in die Studie (*Ausgangswert, T1*) sowie drei und sechs Monate nach Studieneinschluss (*Nachbeobachtung, T2 und T3*) (**Abb. 1**). Die IG erhielt ein interdisziplinäres multimodales Assessment (IMA), die KG eine unimodale Eingangsdiagnostik. Relevant für die Fragestellung der Studie war der Vergleich der beiden Studienarme im Hinblick auf die Veränderungen der zwei primären Zielgrößen im Zeitverlauf.

### 2.1.2 Sekundäre Zielgrößen

Im Folgenden werden zunächst die ausgewerteten sekundären Zielgrößen gemäß des Evaluationskonzeptes beschrieben.

#### Behandlungserfolg aus Patientensicht (Globales Veränderungsitem aus VFB, Frage 13)

Der von den Patienten selbsteingeschätzte Behandlungserfolg wurde mithilfe eines im VFB fest integrierten Einzelitems gemessen („Wenn Sie alles zusammen betrachten, wie beurteilen Sie den bisherigen Erfolg Ihrer Behandlung?“). Als Antwortoption standen Ausprägungen von „sehr gut“ (1) bis „schlecht“ (5) zur Auswahl.

#### Depression-Angst-Stress-Skala (DASS) (Lovibond & Lovibond 1995, Nilges & Essau 2015)

Um insbesondere der Depression als häufig beobachteter Komorbidität von Schmerzen Rechnung zu tragen, legen die insgesamt 21 Fragen des DASS mit jeweils 7 Items pro Subskala den Fokus auf psychologische Kernaspekte von Depression, Angst und Stress. Die Aussagen der Items sind negativ konnotiert (z.B. „Ich fand es schwer, mich zu beruhigen.“) und ermöglichen Antworten zwischen „Traf gar nicht auf mich zu“ (0) und „Traf sehr stark auf mich zu oder die meiste Zeit“ (3). Die Berechnung der Skalenwerte erfolgt durch einfache Aufsummierung der Antwortpunkte, wobei je Subskala ein fehlender Wert zulässig ist und durch den Mittelwert der übrigen Items einer Subskala ersetzt werden kann. Bei den Subskalen Depressivität (D) und Stress (S) wird ab 10 Punkten eine erhöhte Ausprägung angenommen. Bei der Skala Angst (A) besteht ab 6 Punkten eine erhöhte Wahrscheinlichkeit für das Vorliegen einer ausgeprägten Belastung durch eine Angststörung.

#### Gesundheitsbezogene Lebensqualität (VR-12) (Kazis et al. 2004) (Buchholz et al. 2017)

Aus den insgesamt 12 Items des VR-12 mit wechselnden Antwortformaten lässt sich mit spezifischen Item-Gewichtungen ein körperlicher Summenscore (Physical Composite Summary, PCS) und ein psychischer Summenscore (Mental Composite Summary, MCS) berechnen. Beide Summenscores sind so normiert, dass sich in einer Standardpopulation ein Mittelwert von 50 und eine Standardabweichung von 10 ergeben. Über dem Mittelwert liegende Scores sprechen für eine überdurchschnittliche Lebensqualität bzgl. der jeweiligen Komponente. Für die Auswertungen in PAIN2020 wurden eigene SPSS-Syntaxen verwendet, die auf einem durch Kazis persönlich zur Verfügung gestellten SAS-Code basieren. Hierdurch konnten für die Berechnung der Summenscores zwei fehlende Werte zugelassen werden.

Auf die Auswertung der selbstberichteten *Arbeitsunfähigkeitstage* (AU-Tage) aus dem DSF und VFB („Wie viele Tage waren Sie in den letzten 3 Monaten arbeitsunfähig?“) wurde verzichtet, da im DSF für dieses Item in 33,9% der Fälle zu T1 kein Wert vorlag. Es werden statt der selbstberichteten AU-Tage, welche durch potenzielle Erinnerungsfehler zudem einer gewissen Ungenauigkeit unterliegen können, die im Rahmen des Evaluationsansatzes 2 erhobenen AU-Tage aus den Routinedaten des BARMER Wissenschafts-Datawarehouse dargestellt (vgl. Kapitel 2.2.1).

Der zu den drei Messzeitpunkten eingesetzte DSF bzw. seine Verlaufsversion umfassen routinemäßig weitere folgende Erhebungsinstrumente, die als sekundäre Zielgrößen zusätzlich ausgewertet wurden:

### Schmerzbeschreibungsliste (SBL) (Korb & Pfinzgen 2003)

Während die ersten 8 Adjektive der SBL eine qualitative sensorische Schmerzkomponente bilden und u.a. die Differenzialdiagnostik einzelner Schmerzsyndrome (z.B. Kopfschmerz, neuropathischer Schmerz) unterstützen, wird aus den letzten 4 Schmerzbeschreibungen („Ich empfinde meine Schmerzen als...“ „...elend“, „...schauderhaft“, „...scheußlich“, „...furchtbar“) ein affektiver Summenwert gebildet (0-12). Die Antwortmöglichkeiten reichen von „trifft genau zu“ (3) bis „trifft nicht zu“ (0). Werte zwischen 8-12 sprechen für ein erhöhtes affektives Schmerzerleben und gelten als auffällig.

### Habituelles Wohlbefinden (MFHW) (Herda et al. 1998)

Der Marburger Fragebogen zum habituellen Wohlbefinden besteht aus 7 positiv formulierten Items (z.B. „Ich habe mich richtig freuen können.“). Die 6-stufige Antwortskala bewegt sich zwischen „trifft gar nicht zu“ (0) und „trifft vollkommen zu“ (5). Eine einfache Addition der Itemwerte führt zu einem zwischen 0-35 liegenden Summenwert, wobei der Maximalwert ein besonders starkes Wohlbefinden ausdrückt. Zur Berechnung des Summenwertes ist eine fehlende Angabe zulässig, die durch den Mittelwert der restlichen 6 Items ersetzt wird. Ein Summenwert zwischen 0-10 wird für Schmerzpatienten als niedrig und daher auffällig angesehen.

### Pain Catastrophizing Scale (Sullivan 2009, Meyer et al. 2008)

Die Pain-Catastrophizing-Skala wurde in PAIN2020 dem standardisierten Fragebogenset des DSF und des VFB hinzugefügt, um das Katastrophisieren zu messen, welches einen wichtigen Prädiktor für den Schweregrad und die Art der Schmerzbewältigung darstellt. Die insgesamt 13 Items sind negativ besetzt (z.B. „Ich kann nicht aufhören, an die Schmerzen zu denken.“) und geben die Antwortoptionen „trifft überhaupt nicht zu“ (0) bis „trifft immer zu“ (4) vor. Es wird ein gesamter Summenwert berechnet (0-52), der zudem in die Subskalen „Hilflosigkeit“ (0-20), „Verstärkung“ (0-12) und „Grübeln“ (0-20) aufgegliedert wird. Zur Ermittlung des gesamten Summenwertes sind zwei fehlende Werte erlaubt, für die Berechnung der Subskala „Hilflosigkeit“ einer. Für die Subskalen „Verstärkung“ und „Grübeln“ hingegen ist keine Ersetzung fehlender Angaben durch den Mittelwert vorgesehen.

### Patientenzufriedenheit (ZUF-8) (Schmidt et al. 1989)

Der ZUF-8 ist ein kurzer Fragebogen zur Erfassung der allgemeinen bzw. generellen Patientenzufriedenheit, bestehend aus 8 Items mit jeweils vier (wechselnden) Antwortmöglichkeiten (1-4). Er ist originär nicht im VFB enthalten. Der ZUF-8 wird ursprünglich am Ende stationärer Behandlungen in Rehabilitationskliniken oder Krankenhäusern sowie bei stationärer Psychotherapie bzw. psychosomatischer Rehabilitation eingesetzt. In Rücksprache mit den Entwicklern wurde er in PAIN2020 für ein generisches Setting leicht modifiziert. Der nach Umpolung von 4 Items einfach aufsummierte Skalenwert reicht von 8-32, wobei hohe Werte eine hohe Zufriedenheit kennzeichnen.

Der modifizierte ZUF-8 wurde abweichend vom Evaluationskonzept zur Ergänzung der sekundären Zielgröße „Behandlungserfolg aus Patientensicht“ im VFB eingesetzt. Die Patientenzufriedenheit ist ein bedeutender, mehrdimensionaler Outcome-Parameter (Donabedian 1988), welcher der Patientenorientierung in der Krankenversorgung Rechnung trägt (Lecher et al. 2002) und mittels des im VFB routinemäßig enthaltenen Einzelitem zum Behandlungserfolg aus Patientensicht nicht oder nur unzureichend abgebildet wird.

Über die Ergebnisse dieser zusätzlich ausgewerteten Zielgrößen wird in der Ergebnisdarstellung weiter unten (Kapitel 3.1.6) getrennt von den Ergebnissen für die im Evaluationskonzept enthaltenen sekundären Zielgrößen berichtet.

### 2.1.3 Ein- und Ausschlusskriterien

#### Einschlusskriterien

Durch die Studienzentren wurden Teilnehmer eingeschlossen, auf die folgende Merkmale zutreffen mussten:

- Mindestalter 18 Jahre und BARMER-versichert<sup>1</sup>,
- Schmerzen:
  - seit mind. 6 Wochen und/oder
  - in kürzeren Phasen innerhalb der letzten beiden Jahre mehrfach auftretend,
- Schmerzbedingte Einschränkungen im Lebensvollzug:
  - Schmerzbedingte Arbeitsunfähigkeit (AU) seit mind. 4 Wochen oder kumulierte AU von mind. 6 Wochen in den vergangenen 12 Monaten und/oder
  - Schmerzbedingte deutliche Beeinträchtigung in Familie/Freizeit/Arbeit,
- Die Erreichbarkeit des Behandlungszentrums ist gewährleistet.

Zusätzlich musste bei den folgenden aufgeführten Kriterien mindestens eines zutreffend sein:

- Risikofaktoren für Chronifizierung erkennbar:
  - bzgl. Lokalisation: sich ausbreitende Schmerzen,
  - ausgeprägtes (verbales/non-verbales) Schmerzverhalten,
  - ungünstige Schmerzverarbeitung (Fokussierung, Ängste, ...),
  - ausgeprägtes Schon- und Vermeidungsverhalten,
  - Hinweise auf Stressbelastung in Familie/Partnerschaft/soz. Umfeld/Beruf,
  - depressive Symptome im Erleben und/oder Verhalten,
  - Befindlichkeit geprägt durch Frustration/Ärger,
  - ungünstige Verhaltensweisen (Überforderung, „Durchhalten“),
  - Hinweise auf Somatisierung (z.B. vielfältige, „bunte“ Symptomatik),
  - hohes Inanspruchnahmeverhalten im Versorgungssystem,
  - Wunsch nach fortgesetzter Krankschreibung,
  - Wunsch nach fortgesetzter Diagnostik,
- Mindestens 3 positive Angaben im Fragebogen für Patienten mit länger anhaltenden Schmerzen (siehe **Anhang 1**),
- Ungeklärte Fragen (über Diagnose, Pathologie, über erforderlichen Krankenhausaufenthalt oder über Erfordernis operativen Vorgehens, Reha-Bedarf).

#### Ausschlusskriterien

Traf eines oder mehrere der folgenden Kriterien zu, führte dies zum Ausschluss:

- Andere, akut wirksame schwerwiegende Erkrankungen, die z.B. eine aktivierende Behandlung verhindern,
- Eindeutige „Red Flags“,
- Manifeste chronische Schmerzerkrankung (z.B. AU wegen Schmerzen länger als 6 Monate, schmerzrelevante Diagnose über mehr als 4 Quartale, vorhergehende Therapie mit starken Opioiden über 3 Monate Dauer, vorhergehende interdisziplinäre multimodale Schmerztherapie in letzten beiden Jahren; FMS),
- Laufende Diagnostik und/oder Therapie bzgl. schmerzspezifischer Pathologie,

---

<sup>1</sup> Zu einem späteren Zeitpunkt der Studie trat die Kaufmännische Krankenkasse (KKH) dem Selektivvertrag bei.

- Schwere und aktive psychiatrische Störung (Persönlichkeitsstörung, schwere Depression oder Angsterkrankung, Hinweise auf Suizidalität),
- Laufendes Renten- oder Reha-Verfahren (beantragt),
- Sprachliche und/oder kognitive Einschränkungen, die eine Durchführung des IMA bzw. der schmerztherapeutischen Regelversorgung unmöglich erscheinen ließen.

#### 2.1.4 Berechnung der Stichprobengröße

Die Berechnung der erforderlichen Stichprobengröße basierte auf den im Evaluationsansatz 1 zur Untersuchung der beiden primären Zielgrößen (SI, SBF) geplanten Analysen. Ziel war es, mit der geplanten Stichprobengröße Gruppenunterschiede im Bereich einer standardisierten Mittelwertdifferenz  $d$  von  $\geq 0,10$  mit einer Power von 80 % statistisch abzusichern. Den Berechnungen lag eine Anzahl von 29 teilnehmenden Einrichtungen zugrunde.

Zunächst wurde festgelegt, dass bei der Irrtumswahrscheinlichkeit für den Fehler 1. Art in einem zweiseitigen Test ein Signifikanzniveau von  $\alpha = 0,05$  gelten soll. Die Wahrscheinlichkeit, dass bei Vorliegen eines Gruppenunterschiedes dieser auch statistisch erkannt wird, soll  $1 - \beta = 0,80$  betragen. Zur Korrektur für multiples Testen (zwei Zielgrößen, zwei Messzeitpunkte) wird der  $\alpha$  - Fehler nach Bonferroni auf  $\alpha = 0,05/4 = 0,0125$  adjustiert.

Fallzahlberechnungen für statistische Analysen von Mehrebenenendaten (Zentren, Patienten, Zeitpunkte) werden durch drei Sachverhalte erschwert: Es müssen erstens im Unterschied zu Analysen mit weniger komplexen Daten die Ausprägungen zahlreicher Modellparameter vorgegeben werden (z.B. für die Varianzkomponenten), bei denen zweitens häufig eine gewisse Unsicherheit über realistische Werte oder Wertbereiche besteht (z.B. für die Größe fester Effekte auf den verschiedenen Ebenen). Drittens ist die Berechnung von Fallzahlen oder der Power bei Mehrebenenmodellen im Allgemeinen nicht in geschlossener Form möglich, so dass die erreichte Power nach Vorgabe von Modellparametern in Simulationsrechnungen bestimmt werden muss.

Vor diesem Hintergrund wurden die Power- und Fallzahlbetrachtungen zwar anhand eines Mehrebenenmodells durchgeführt, es wurde aber ein vereinfachtes Modell betrachtet, bei dem Gruppenunterschiede zwischen standardisierten Mittelwertdifferenzen (SRMs) für die zwei primären Zielgrößen untersucht werden. Die SRMs werden in diesem Modell durch Division der Prä-Post-Unterschiede zwischen Ausgangswert- und Nachbeobachtung durch die jeweilige Standardabweichung der Prä-Post-Unterschiede berechnet ( $SRM_{t1-t2}$ ,  $SRM_{t1-t3}$ ). Gruppenunterschiede werden in Varianzanalysen mit den SRMs als abhängiger Variable und der Gruppenzugehörigkeit als Prädiktor statistisch getestet. Dies entspricht dem Test eines Interaktionseffekts Gruppe\*Zeit in einer Varianzanalyse mit Messwiederholung (zwei Messzeitpunkte) und einer dichotomen Gruppenvariable.

Die Simulation der Power in diesem gegenüber dem weiter oben formulierten Grundmodell vereinfachten Modell erfolgte mit dem Programm MLPowSim (Browne et al. 2009). Dabei wurden die in **Tab. 1** aufgeführten technischen Parameter zugrunde gelegt. Als Effektgrößen wurden auf die Residualvarianz standardisierte Unterschiede zwischen den SRMs verwendet. Es wurde dabei festgelegt, dass die Residualvarianz in den Simulationsrechnungen 1.0 beträgt. Die Unterschiede zwischen den Mittelwerten können dadurch direkt als Effektgröße interpretiert werden.

Bei der Wahl der in der Fallzahlplanung zu berücksichtigenden Effektgrößen wurde davon ausgegangen, dass diese in einem Bereich lokalisiert sind, der eher kleine Effekte beschreibt. Diese Annahme war durch mehrere Faktoren begründet: Die in der neuen Versorgungsform zu evaluierende Intervention beinhaltete primär ein multimodales Assessment. Die Ergebnisse des Assessments sollten zwar verschiedene Therapieoptionen ansteuern, es war aber zu erwarten, dass bei dieser Steuerung in den verschiedenen Zentren eine gewisse Variabilität

existieren wird, durch die die beobachteten Effekte verdünnt werden könnten. Darüber hinaus war zu berücksichtigen, dass die Art der Zuweisung der Patienten vermutlich eine weitere Quelle von Heterogenität sein würde. Bei den zugewiesenen Patienten würde es sich um eine eher uneinheitliche Gruppe handeln, die durch ein weites Spektrum von krankheits- und schmerzbezogenen Merkmalen charakterisiert ist. Das multimodale Assessment und die darauf aufbauende Behandlungssteuerung würde nicht für alle Untergruppen gleich große Effekte ergeben können. Auch dadurch könnten dann Effekte verdünnt werden.

Wegen der Annahme kleiner Effekte wurden die Simulationsrechnungen für vier Effektgrößen im Wertebereich von 0,10 bis 0,20 durchgeführt. Zur Überprüfung des Einflusses verschiedener Szenarien der Verteilung der Zufallseffekte auf die Ebene der Patienten bzw. der Zentren wurden weiterhin drei Szenarien mit einer Verteilung im Wertebereich von 1,0 : 0 bis 0,8 : 0,2 (1. Ebene : 2. Ebene) einbezogen. In den Simulationsrechnungen wurde die Fallzahl in den einzelnen Zentren systematisch variiert, bis bei 1.000 Simulationen pro Fallzahl-Szenario eine Power von 80 % erreicht wurde.

Tab. 1: Technische Parameter der Simulationsrechnungen

Art des Modells	Zwei-Ebenen-Modell mit Zufallseffekten (Random Intercept)
Alpha-Niveau (zweiseitiger Test)	Nominell $\alpha = 0,05$ mit Bonferroni-Adjustierung ( $\alpha = 0,00125$ )
Simulationen je Fallzahl-Szenario	1.000
Berechnung der Modelle	R 2.15; Bibliotheken MASS, lme4
Schätzalgorithmus in R	IWLS
Gruppenvariable (fester Effekt)	Normalverteilungsapproximation*
Effektgröße**	0,10 / 0,15 / 0,175 / 0,20
Varianzkomponenten 1. : 2. Ebene	1,0 : 0,0 / 0,9 : 0,1 / 0,8 : 0,2

\* mit Gruppenverhältnis 70 : 30 (Interventionsgruppe : Kontrollgruppe)

\*\* in Einheiten der Residualvarianz

Nach den Ergebnissen der Simulationsrechnungen mit MLPowSim (**Tab. 2**) verlief die Fallzahlkurve in Abhängigkeit von den vorgegebenen Effektgrößen sehr steil, während das Verhältnis der Varianzkomponenten einen weitaus geringeren Einfluss auf die benötigte Stichprobengröße hatte. Nach diesen Szenarien würde eine Stichprobengröße im Bereich von etwa 2.000 Fällen unter den gegebenen Annahmen (z.B. Effektgröße 0,175; Verteilung der Varianzkomponenten 0,9 : 0,1) ausreichen, um die primäre Fragestellung zu beantworten.

Tab. 2: Gesamt-Fallzahlen in Zwei-Ebenen-Modellen, bei denen die Power 80 % beträgt

Varianzkomponenten (1. Ebene : 2. Ebene)	Effektgröße			
	0,10	0,15	0,175	0,20
1,0 : 0,0	5.300	2.360	1.730	1.340
0,9 : 0,1	5.920	2.670	1.980	1.510
0,8 : 0,2	6.670	2.990	2.210	1.690

Bei der Beurteilung dieses Ergebnisses war jedoch zu berücksichtigen, dass dieses auf vereinfachten Annahmen beruht. So wurde bei den Simulationen davon ausgegangen, dass in jedem der 29 Zentren jeweils die gleiche Anzahl von Fällen rekrutiert wird. Abweichungen von dieser (eher unrealistischen) Annahme dürften zu einem höheren notwendigen Stichprobenumfang führen. Daneben wurde angenommen, dass über die auf die Regressionskonstante bezogenen Zufallseffekte hinaus keine weiteren Zufallseffekte von Bedeutung sind. In hier nicht im Einzelnen gezeigten ergänzenden Analysen ergab sich, dass bei Hinzunahme eines auf den Gruppenunterschied bezogenen Zufallseffekts unter realistischen Annahmen etwa 10 % mehr Fälle benötigt werden. Nicht zuletzt ergab die im



Bereich unterhalb der Effektgröße 0,175 steile Fallzahlkurve, dass auch hier zur Absicherung gegen geringfügig kleinere Effekte eine etwas höhere Fallzahl notwendig ist. Diese Überlegung wurde auch durch Ergebnisse aus systematischen Übersichten und Meta-Analysen gestützt, bei denen z.B. beim Vergleich konservativer Behandlungsoptionen bei chronischen Rückenschmerzen in einigen Fällen für die Endpunkte Schmerz und Funktionseinschränkungen aggregierte Mittelwertunterschiede errechnet wurden, die merklich unter dem Wert von  $d = 0,2$  liegen (O’Keeffe et al. 2016).

Nach den Erfahrungen in ähnlichen Vorhaben war mit einem nicht zu vernachlässigenden Probandenverlust (Drop-Out) im Verlauf der Studie zu rechnen. Es wurde erwartet, dass von den zur Studienteilnahme bereiten Patienten etwa jeweils 30 % beim Übergang zwischen den Messzeitpunkten aus der Studie ausscheiden. Von den für die Studie geeigneten Patienten würden sich etwa 70 % für eine Studienteilnahme entscheiden. In **Tab. 3** werden die Brutto-Fallzahlen gezeigt, die sich aus der anhand dieser Annahmen durchgeführten Rückberechnung ergaben. Als Netto-Fallzahl wurde die in der Power-Simulation als realistisch eingeschätzte Gesamt-Fallzahl von 2.200 Patienten (als Annäherung an die bei 29 teilnehmenden Einrichtungen erwarteten 2.175 Fälle) zugrunde gelegt. Damit ergab sich bei moderater Erhöhung der in der Simulationsrechnung erhaltenen Netto-Stichprobengröße eine initiale Bruttostichprobe im Bereich von 4.500 Patienten.

Tab. 3: Rückberechnung der Brutto-Fallzahlen bei Berücksichtigung von Drop-Outs

<b>Probandenverlust</b>		
Fallzahl T3 (70 % von T2)	Fallzahl T2 (70 % von T1)	Fallzahl T1 100 %
2.205	3.150	4.500

### 2.1.5 Auswertungsvarianten

Entsprechend des Evaluationsansatzes 1 lag der Fokus bei der Datenauswertung auf der Identifizierung von Unterschieden zwischen der IG und der KG im Nachbeobachtungszeitraum. Hierbei fand das Intention-to-treat-Prinzip (ITT) Anwendung, bei dem alle Daten von Teilnehmenden auch dann in die Auswertung einbezogen werden, wenn diese nach Einschluss vorzeitig aus der Studie ausscheiden oder die Therapie wechseln. Im Kontext der vorliegenden Studie bedeutet dies, es wurden für die Auswertung alle Teilnehmenden berücksichtigt, die sowohl randomisiert wurden als auch ein Assessment erhalten hatten.

Darüber hinaus wurden zur Beantwortung der Fragestellung im Evaluationsansatz I drei Auswertungsvarianten gewählt:

#### Auswertungsvariante I: Zufallskoeffizientenmodelle

Wie im Evaluationskonzept ausgeführt, wurden in einer ersten Auswertungsvariante Regressionsmodelle berechnet, in denen alle über die drei Messzeitpunkte verfügbaren Daten der primären und sekundären Zielgrößen in die Berechnung aufgenommen wurden. Den Ergebnissen liegt ein Regressionsmodell für Messwiederholungsdaten und gemischte Effekte zugrunde (mixed effects model for repeated measures, MMRM). Die Werteausprägungen der primären und sekundären Zielgrößen an den Messzeitpunkten T2 und T3 wurden aus a) den Ausgangswerten der Zielgrößen zum ersten Messzeitpunkt ( $y_{t1}$ ), b) den Faktoren der Gruppenzugehörigkeit (*Gruppe*) und des betreffenden Zeitpunkts (*Zeit*) sowie c) dem Interaktionseffekt von Gruppenzugehörigkeit und Zeitpunkt vorhergesagt. Für die Regressionskonstante wurde ein Zufallseffekt in das Modell aufgenommen.

$$y_t = a + b_1 * y_{t1} + b_2 * \text{Gruppe} + b_3 * \text{Zeit} + b_4 * \text{Gruppe} \times \text{Zeit}$$

Da Ausgangswertunterschiede in diesen Modellen bereits berücksichtigt sind, interessieren insbesondere die Effekte der Gruppenzugehörigkeit, da sie Unterschiede in den Verläufen beschreiben. Der Zeiteffekt markiert hingegen, ob es unabhängig von der Gruppenzugehörigkeit Veränderungen von T1 zu T2 bzw. T3 gegeben hat. Der Interaktionseffekt beschreibt Unterschiede in diesen Veränderungen in Abhängigkeit vom betrachteten Messzeitpunkt.

Die Regressionsmodelle wurden so parametrisiert, dass eine Zentrierung der Ausgangswerte (T1) erfolgte und die Kontrollgruppe bzw. der Messzeitpunkt T3 als Referenzkategorie definiert wurden. Die Regressionskonstante beschreibt deshalb den unter dem Modell erwarteten Wert der jeweiligen abhängigen Variable in der Kontrollgruppe zum Messzeitpunkt T3. Der Zeiteffekt beschreibt, in welchem Maße der Wert zu T2 von dem zu T3 abweicht. Der Gruppeneffekt beschreibt, in welchem Maße der Wert der Interventionsgruppe von dem der Kontrollgruppe abweicht. Analoges gilt für den Interaktionseffekt. Das Auswertungsmodell entspricht damit einer Kovarianzanalyse, wie sie für die Analyse von kontrollierten Studien mit Längsschnittdaten als Alternative zur Verwendung von Differenzwerten zur Kontrolle von Ausgangswerten empfohlen wird (Vickers und Altman 2001).

#### Auswertungsvariante II: Complete-Case-Analyse

Zur Absicherung der Ergebnisse aus den Zufallskoeffizientenmodellen wurden in einer zweiten Auswertungsvariante nur Datensätze mit Werten zu allen drei Messzeitpunkten berücksichtigt. Schwerpunkt dieses Complete-Case-Ansatzes bildeten Varianzanalysen mit Messwiederholungen zur Identifikation von Interaktionseffekten zwischen der jeweiligen primären oder sekundären Zielgröße und der Gruppenzugehörigkeit (IG, KG). Effektgrößen zwischen zwei Messzeitpunkten wurden mit Cohens  $d$ , über alle drei Messzeitpunkte mit Cohens  $f$  abgebildet.

#### Auswertungsvariante III: Available-Case-Analyse

In einer dritten Auswertungsvariante wurden für deskriptive Ergebnisdarstellungen und Gruppentestungen (t-Tests) sowie die Effektgröße Cohens  $d$  alle Fälle einbezogen, die für die Datenanalyse zu den drei Messzeitpunkten T1, T2 und T3 *jeweils* zur Verfügung standen. Auf eine Imputation von fehlenden Werten wurde verzichtet.

Für alle drei Varianten erfolgten die Auswertungen mit IBM SPSS Statistics 28.

### **2.1.6 Qualitätssicherungsmaßnahmen**

Unabhängig von eigenen Qualitätssicherungsinstrumenten der PAIN2020-Konsortialpartner wurde die Qualität der in die studieneigene Datenbank eingehenden Primärdaten durch das Evaluationsinstitut regelmäßig monitoriert und die Ergebnisse dem Konsortium zurückgemeldet. Maßnahmen zur Qualitätssicherung umfassten insbesondere:

- die Prüfung des Primärdatensatzes auf Vollständigkeit, Plausibilität und fehlende Werte,
- die Reproduktion von zur Verfügung gestellten Score-Werten,
- stichprobenhafte Eingabekontrollen (insbesondere der für die Zielgrößen relevanten Dokumente),
- Unterstützung bei der Entwicklung der Anschreiben für die Verlaufsbefragung,
- Unterstützung bei der Entwicklung der online-gestützten Verlaufsbefragung,
- die regelmäßige Rückmeldung von Rücklaufquoten (auch zentrumsspezifischer) und hierauf basierende Fallzahlextrapolationen.

## 2.2 Evaluationsansatz 2

Gegenstand der Analysen im Evaluationsansatz 2 im Sinne einer gesundheitsökonomischen Evaluation war die vergleichende Untersuchung von Sekundärdaten der BARMER Krankenkasse für die im Evaluationsansatz 1 eingeschlossenen Studienteilnehmer aus Interventions- und Kontrollgruppe. Die Leistungsdaten standen durch Ferndatenverarbeitung im Wissenschafts-Datwarehouse der BARMER (W-DWH) zur Verfügung.

### 2.2.1 Sekundäre Zielgrößen

Für Evaluationsansatz 2 wurden folgende Leistungsbereiche des W-DWH berücksichtigt:

- *Ambulante vertragsärztliche Versorgung (AMB),*
- *Ambulante Krankenhausleistungen (KHAMB),*
- *Stationäre Krankenhausleistungen (KH),*
- *Arzneimittel (AM),*
- *Heil- und Hilfsmittel (HEILHILF),*
- *Stationäre Rehabilitationsleistungen (REHA),*
- *Arbeitsunfähigkeitstage (AU),*
- *Entgeltersatzleistungen (Krankengeld, KG).*

Da Angaben zur *Frühberentung* im W-DWH nicht vorlagen, konnte diese ursprünglich im Evaluationskonzept aufgeführte sekundäre Zielgröße nicht ausgewertet werden.

Arbeitsunfähigkeitstage und Krankengeld ausgenommen, ergeben die einzelnen Kostenpositionen in der Addition die Summe der direkten (medizinischen) Kosten (Greiner & Schöffski 2012). Mit Ausnahme der Arbeitsunfähigkeitstage wurden die Leistungen jeweils in **Euro**-Beträgen abgebildet. In der Regel lagen die Daten in dieser Einheit in den verschiedenen Leistungsbibliotheken des W-DWH vor. Für die ambulante vertragsärztliche Versorgung ergab sich durch die Abrechnung auf der Grundlage des Einheitlichen Bewertungsmaßstabes allerdings die Besonderheit, dass Leistungen lediglich ein Punktwert zugewiesen sein kann. In einem solchen Fall wurde die Punktzahl mit dem jeweiligen Orientierungspunktwert aus dem Jahr, in dem die Leistung abgerechnet wurde, multipliziert (Neubauer et al. 2014).

Die direkten Kosten, Arbeitsunfähigkeitstage und das Krankengeld geben die *Perspektive der gesetzlichen Krankenversicherung* wieder, das heißt, Zuzahlungen (z.B. zu Arznei- oder Heilmitteln) oder etwa private Transportkosten bleiben unberücksichtigt (Greiner & Schöffski 2012).

Leistungen im Rahmen der PAIN2020-Studie wurden mit spezifischen (BARMER-internen) EBM-Ziffern abgerechnet, sodass diese als *Programmkosten* identifiziert werden konnten.

### 2.2.2 Ein- und Ausschlusskriterien

Ziel im Evaluationsansatz 2 war es, die Kosten und AU-Tage der IG und der KG **sechs Monate vor** sowie **sechs Monate nach dem jeweiligen individuellen Assessment-Termin** zu ermitteln. In die Sekundärdatenanalyse konnten alle Studienteilnehmer aus PAIN2020 einbezogen werden, die:

- für die Auswertung im Evaluationsansatz 1 die Ausgangsstichprobe zu T1 bildeten,
- über den insgesamt einjährigen Beobachtungszeitraum durchgängig bei der BARMER versichert waren,

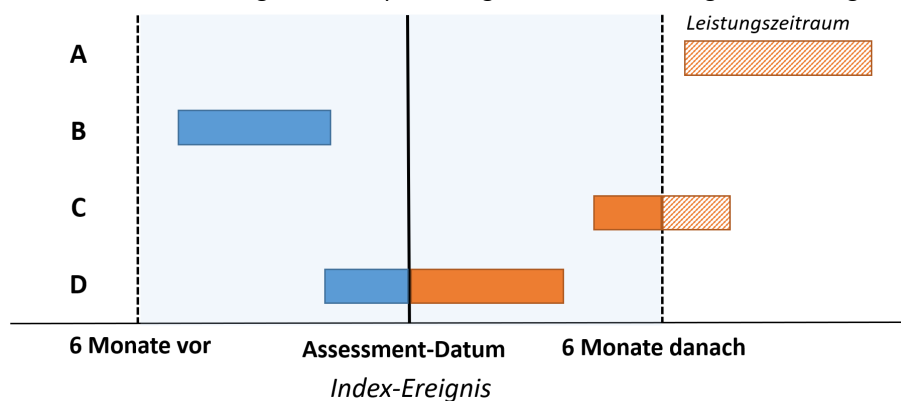


- deren Assessment-Termin spätestens am 30.06.2021 stattfand<sup>2</sup> und
- die im W-DWH zugeordnet werden konnten.

### Periodengerechte Zuordnung der Leistungsdaten

Da Leistungen nicht immer punktuell an einem konkreten Datum anfallen bzw. zwischen Leistungsbeginn und Leistungsende mehrere Tage liegen können (im Folgenden *Leistungszeitraum* genannt), ergaben sich Herausforderungen für die periodengerechte Ermittlung der Kosten und AU-Tage. In **Abb. 2** sind die grundsätzlichen Fallkonstellationen von Leistungszeiträumen bezogen auf den individuellen Beobachtungszeitraum eines Studienteilnehmers dargestellt.

Abb. 2: Herausforderungen bei der periodengerechten Zuordnung von Leistungsdaten



Während die Leistungen in den Fällen A und B entweder innerhalb oder außerhalb des Beobachtungszeitraumes anfallen, überlappen sie in den Fällen C und D dessen Intervallgrenzen. Für die Konstellationen C und D wurden daher im Rahmen der hier erfolgten Sekundärdatenauswertung nur die anteiligen Kosten und AU-Tage eingeschlossen, die dem Beobachtungszeitraum tatsächlich zuzuschreiben waren. Lag der Assessment-Termin wie in Fall D gar innerhalb eines Leistungszeitraumes, wurde der Abstand des Leistungsbeginns bzw. des Leistungsendes zum Assessment-Datum ermittelt und entsprechend des Abstandsverhältnisses ein Teil der Leistung dem 6-Monats-Vorbeobachtungszeitraum und der andere Teil der Leistung dem 6-Monats-Nachbeobachtungszeitraum zugeschrieben.

Im Falle des Krankengeldes kam es – zumindest für die eingeschlossene Stichprobe – bei etwa einem Drittel aller Leistungsfälle in dieser Leistungsbibliothek vor, dass der Beginn oder aber das Ende der Leistung einen fehlenden Wert aufwies, allerdings eine Information über die Anzahl der Leistungstage des gezahlten Krankengeldes verfügbar war. Ein fehlender Beginn einer Leistung ließ sich mithin mittels des Leistungsendes und der Subtraktion der angegebenen Leistungstage berechnen; ein fehlendes Leistungsende analog durch Addition der Leistungstage auf den Leistungsbeginn.

### 2.2.3 Qualitätssicherungsmaßnahmen

Die Annäherung an die komplexe Systematik der verschiedenen zur Verfügung stehenden Leistungsbibliotheken und ihrer sich daraus ergebenden methodischen Herausforderungen (Swart et al. 2015) geschah in enger Zusammenarbeit mit den Administratoren des W-DWHs der BARMER.

<sup>2</sup> Ging der Assessment-Termin über den 30.06.2021 hinaus, so reichte der Nachbeobachtungszeitraum in das Kalenderjahr 2022 hinein. Die Leistungsdaten für ein Kalenderjahr stehen jedoch immer erst im Laufe des Folgejahres vollständig zur Verfügung.

Der Zugriff auf das W-DWH erfolgte per Ferndatenverarbeitung. Studienteilnehmer aus PAIN2020 waren für das Evaluationsinstitut über eine projektinterne ID-Schlüsselliste im W-DWH identifizierbar. Die Plausibilität der extern durch die BARMER per Versichertennummer vorgenommenen Zuordnungen wurde durch einen Abgleich der Merkmale Alter und Geschlecht, die für jeden Studienteilnehmer sowohl durch die primäre Datenerhebung als auch durch die Sekundärdaten vorlagen, systematisch überprüft. Stichprobenhaft wurden weitere Merkmale wie Diagnosen oder die Region (auf 3-stelliger Postleitzahlebene) für die Plausibilitätsüberprüfungen herangezogen.

Aufgrund des hohen Komplexitätsgrades wurden die Syntaxen zur Bestimmung der anteiligen Kosten und AU-Tage für die zwei Beobachtungszeiträume (6 Monate vor sowie 6 Monate nach dem Assessment) im Sinne eines Vieraugenprinzips entwickelt.

#### **2.2.4 Berechnung der Stichprobengröße**

Da im Evaluationsansatz 2 das Problem des Probandenverlustes eine untergeordnete Rolle spielt (bzw. zumindest kein Drop-Out durch fehlenden Rücklauf von Fragebögen wie bei der Primärdatenerhebung entsteht), war davon auszugehen, dass für den Evaluationsansatz 2 mindestens jene für den Evaluationsansatz 1 kalkulierte Gesamt-Nettostichprobe für die Datenauswertung zur Verfügung stünde (N = 2.205 Fälle).

Im Rahmen der Analysen im Evaluationsansatz 2 wurden Gruppenvergleiche (IG vs. KG) durchgeführt. Diese Vergleiche bezogen sich auf die Zeitperioden 6 Monate vor bzw. 6 Monate nach dem Assessment. Gegenüber den Analysen im Evaluationsansatz 1 reduzierte sich dadurch die Komplexität des statistischen Modells in erheblichem Maße. Für die Analyse der Kosten und AU-Tage als stetige Zielgrößen wurde mittels des Programm G\*Power 3.1 unter Zugrundelegung der kalkulierten Gesamt-Nettostichprobe aus Evaluationsansatz 1, des Mann-Whitney-Tests als statistischer Analyseverfahren und einer vorgegebenen Power von 80 % für entdeckbare Gruppenunterschiede für den Vergleich der IG und der KG mit einem normierten Mittelwertsunterschied in Höhe von Cohens  $d = 0,13$  gerechnet.

#### **2.2.5 Auswertungsvarianten**

##### Auswertungsvariante I: Gesamtkosten

Es wurden in der ersten Auswertungsvariante zunächst *alle* für die Studienteilnehmer für die beiden Beobachtungszeiträume (6 Monate vor sowie 6 Monate nach dem Assessment) jeweils auflaufenden Kosten und AU-Tage ermittelt.

##### Auswertungsvariante II: Schmerzbezogene Kosten

Auch wenn es sich bei der eingeschlossenen Studienpopulation um Personen mit Schmerzproblemen handelte, konnten für die Studienteilnehmer Abrechnungsdaten in den Leistungsbibliotheken dokumentiert sein, die nicht mit Schmerzen im Sinne der Einschlusskriterien von PAIN2020 assoziiert waren (Akute Infektionen der oberen Atemwege, Hypertonie, Katarakt etc.). Die in Auswertungsvariante I ausgewiesenen Kosten und AU-Tage haben somit keinen Indikationsbezug. Da die neue Versorgungsform sich jedoch zum Hauptziel setzte, die Versorgungsqualität und Versorgungseffizienz von Patienten mit chronifizierenden Schmerzen zu verbessern, bestand die Möglichkeit, dass potenzielle, durch PAIN2020 indizierte Effekte auf die Inanspruchnahme von Leistungen durch Komorbiditäten überdeckt werden könnten. Vor diesem Hintergrund wurde für die Auswertungsvariante II hinsichtlich der Klassifikationssysteme *ICD-10-Code*, *ATC-Code*, *Heilmittelkatalog* sowie *Hilfsmittelverzeichnis (Anhang 2)* durch das schmerzmedizinische Expertenteam aus dem Kreis der Konsortialpartner von PAIN2020 ein Katalog für indikationsspezifische Einschlusskriterien zur Identifizierung *schmerzbezogener Kosten und AU-Tage* entwickelt (Zeidler et al. 2013).

Da Kostendaten bekanntermaßen sehr schief verteilt und von starken Ausreißern geprägt sein können, fand zur Testung von Gruppenunterschieden zwischen IG und KG der Mann-Whitney-Test (U-Test) Anwendung. Zur Einschätzung der Effekte wurde aus der z-Statistik  $R^2$  berechnet (Fritz et al. 2012).

Für die Auswertung der Sekundärdaten wurden SAS Enterprise Guide 8.3 sowie IBM SPSS Statistics 28 genutzt.

## **2.2.6 Matching einer weiteren Kontrollgruppe aus dem WDWH der BARMER**

Das Evaluationskonzept in PAIN2020 sah für den Evaluationsansatz 2 darüber hinaus einen ergänzenden Vergleich der Routinedaten der IG mit den Routinedaten einer aus dem Gesamtversichertenbestand der BARMER gewonnenen weiteren KG vor. Die Auswahl geeigneter Fälle aus dem Gesamtdatenbestand sollte mit Hilfe des Propensity-Score-Matching erfolgen. Dadurch sollte sichergestellt werden, dass aus dem Gesamtdatenbestand solche Versicherte ausgewählt werden, die sich in relevanten Merkmalen möglichst wenig von den Studienpatienten in der IG unterscheiden.

Für die Bestimmung der Kosten und der AU-Tage für eine gematchte KG aus dem Versichertenbestand der BARMER wurden folgende Arbeitsschritte durchgeführt:

- 1) Vorabselektion einer Stichprobe mit potenziellen Matchingpartnern
- 2) Durchführung eines um Elemente des exakten Matchings erweiterten Propensity-Score-Matchings
- 3) Vergleich der Durchschnittskosten von IG und Matchingpartnern in der Nachbeobachtung

### zu 1) Vorabselektion einer Stichprobe potenzieller Matchingpartner

Das via Ferndatenverarbeitung erreichbare WDWH der BARMER umfasst ca. 8,7 Mio. Versicherte. Ein Leistungsbereich kann eine Vielzahl an Einträgen für einen Versicherten enthalten, wenn dieser innerhalb eines Kalenderjahres z.B. einen Arzt mehrmals aufsuchte, unterschiedliche (Fach-)Ärzte aufsuchte, Komorbiditäten vorliegen, mehrere Arzneimittel verordnet bekam, wiederholt arbeitsunfähig krankgeschrieben wurde usw. Da das Arbeiten mit Datensätzen dieser Größenordnung (u.a. Abfragen, Zusammenfügen von Datentabellen, umfangreichere statistische Prozeduren) technisch sehr ressourcenintensiv ist und mit sehr langen Rechenzeiten verbunden sein kann, wurde als erster Arbeitsschritt eine Vorabselektion zur Verkleinerung der Stichprobe potenzieller Matchingpartner vorgenommen. So wurden im WDWH zunächst Versicherte mit folgenden Eigenschaften gesucht:

- Alter ab 18 Jahren,
- keine Teilnahme an PAIN2020,
- Auftreten mindestens einer der vom Projektteam festgelegten schmerzbezogenen ICD-Diagnosen im ambulanten vertragsärztlichen Bereich im Zeitraum zwischen 2018-2021,
- durchgängige Versicherung bei der BARMER im Zeitraum zwischen 2018-2021,
- erste zwei Ziffern der Postleitzahl des Wohnortes innerhalb oder zumindest im unmittelbaren Umland des Postleitzahlbereiches eines in PAIN2020 beteiligten Schmerzzentrums.

Die auf diese Weise definierte Stichprobe umfasste 2,8 Mio. Versicherte.

### zu 2) Durchführung des Propensity-Score-Matchings

Ziel des Propensity-Score-Matchings war es, für jeden Teilnehmer der IG (für den Routinedaten vorlagen) einen „Zwilling“ mit ähnlichen Merkmalseigenschaften im

Versichertenbestand der BARMER zu finden. Das Evaluationskonzept sah folgende Merkmale vor:

- ICD-Diagnosen
- Arbeitsunfähigkeit
- Kostendaten
- Alter
- Geschlecht
- Versicherungsart
- Stellung im Beruf
- Schulbildung
- Region
- Ortsgrößenklasse

Da die Merkmale *Stellung im Beruf*, *Schulbildung* und *Ortsgrößenklasse* nicht im WDWH verfügbar waren, konnten sie nicht für das Matching berücksichtigt werden. Verzichtet wurde ebenfalls auf die Berücksichtigung von Entgeltersatzleistungen, wie z.B. das Krankengeld, da dies u.a. von den Einkommensverhältnissen der Versicherten abhängig ist und einigen kassenspezifischen Besonderheiten unterliegt. Darüber hinaus konnte die ursprüngliche Planung, Postleitzahlbereiche als Merkmal der regionalen Verortung einzubeziehen, nicht realisiert werden, da dieses Merkmal eine zu starke Granularität aufwies, nur als ein exaktes Matching-Kriterium sinnvoll erschien und deshalb zu technischen Problemen bei der Matching-Prozedur führte. Alternativ wurde deshalb für das Merkmal *Region* auf das Bundesland zurückgegriffen.

Als zentrale Kriterien für das Propensity-Score-Matching wurden die Kosten in insgesamt sechs Leistungsbereichen sowie die AU-Tage verwendet. Weitere Prädiktoren waren das Alter, das Geschlecht, der Versicherungsstatus („normal“ versichert, familienversichert), sowie das Auftreten von fünf definierten ICD-Schmerzregionen im Bereich der ambulanten Kosten im PAIN2020-Erhebungszeitraum. Alle genannten Merkmale wurden (mit Ausnahme des Bundeslands) simultan als Prädiktoren der Gruppenzugehörigkeit im Propensity-Score-Matching berücksichtigt. Darüber hinaus wurden die Merkmale *Geschlecht* und *Bundesland* als exakte Matchingkriterien definiert, d.h. ein Matchingpartner wurde nur dann einem nach dem Propensity Score ähnlichen Partner zugeordnet, wenn dieser dasselbe Geschlecht hatte und aus demselben Bundesland stammte.

Eine besondere Herausforderung bildete die Ermittlung der schmerzbezogenen *Kostendaten*. Die bisher auf diesem Gebiet durchgeführten Analysen orientierten sich jeweils an dem individuellen Assessment-Termin eines PAIN2020-Teilnehmers in der IG oder KG, wodurch sich ebenso individuelle sechsmonatige Vor- und Nachbeobachtungszeiträume ergaben. Für die potenziellen Matchingpartner aus dem Gesamtbestand der BARMER ist ein solcher Assessment-Termin jedoch nicht definiert. Um dennoch geeignete Matchingpartner für PAIN2020-Studienpatienten zu finden, wurde die folgende Vorgehensweise gewählt:

- Für jeden Patienten in der IG wurde das Quartal bestimmt, in dem das Assessment durchgeführt wurde.
- Sodann wurden die im WDWH dokumentierten Kosten in den beiden diesem Quartal vorausgehenden Quartalen getrennt für alle sechs Leistungsbereiche berechnet. Lag der Assessmenttermin beispielsweise im 1. Quartal 2020, dann wurden die Kosten in den beiden „Vorgängerquartalen“ 3. und 4. Quartal 2019 bestimmt.
- Für die Versicherten aus dem Gesamtbestand der BARMER nach Vorabselektion wie unter 1) beschrieben wurden die Kosten in den sechs Leistungsbereichen für alle potenziell in Frage kommenden Vorgängerquartale berechnet.
- Das Matching erfolgte nun derart, dass für einen Patienten der PAIN2020-Interventionsgruppe im Matching-Verfahren nur die Kostendaten aus dem

Gesamtbestand der BARMER herangezogen wurden, welche die entsprechenden beiden Vorgängerquartale dieses Patienten betrafen.

Tab. 4: Realisiertes Matchingmodell

<b>Evaluationskonzept</b>	<b>Matchinganalyse</b>	<b>Kodierung</b>	<b>Matchingverfahren</b>
<b>Soziodemographie</b>			
Alter	Alter	in Jahren	PS
Geschlecht	Geschlecht	0 = männlich 1 = weiblich	PS, exakt
Versicherungsart	Familienversichert	0 = nein 1 = ja	PS
Region	Bundesland	01 - 16 (nach statistischem Bundesamt)	exakt
Ortsgrößenklasse	Nicht realisiert		
Schulbildung	Nicht realisiert		
Stellung im Beruf	Nicht realisiert		
<b>Schmerz</b>			
ICD-Diagnosen	Gruppen: Kopfschmerz Gelenkschmerz Schmerzen Rücken / Wirbelsäule Schmerzen Weichteilgewebe Sonstige Schmerzen	0 = kein Auftreten 1 = Auftreten (im PAIN2020- Erhebungszeitraum)	PS
<b>Leistungsbereich</b>			
Kosten	AMB KHAMB KH AM HEILHILF REHA	Kosten in €	PS
		zusätzlich jeweils gruppiert	exakt
Arbeitsunfähigkeit	AU	Tage	PS
		zusätzlich gruppiert	exakt

AMB: Ambulante vertragsärztliche Versorgung, KHAMB: Ambulante Krankenhausleistungen, KH: Stationäre Krankenhausleistungen, AM: Arzneimittel, HEILHILF: Heil- und Hilfsmittel, REHA: Stationäre Rehabilitationsleistungen, AU: Arbeitsunfähigkeitstage, PS: Das Merkmal wurde für den Propensity-Score berücksichtigt.

Wie die Analysen zeigten, konnte mit dem Propensity-Score-Matching auf dieser Basis keine ausreichend gut vergleichbare Gruppe aus dem Gesamtbestand gebildet werden. Besonders im Bereich der Kostendaten traten zwischen der PAIN2020-Interventionsgruppe und der Vergleichsgruppe aus dem Gesamtbestand der BARMER relevante Unterschiede auf. Aus diesem Grunde wurde das Propensity-Score-Matching erweitert, indem zusätzliche Bedingungen des exakten Matchings eingeführt wurden. Hierfür wurden für alle Kostenbereiche und für die AU-Tage die Merkmalsausprägungen in drei oder vier Kategorien gruppiert (1: keine Kosten, 2: Kosten >0 bis 100€, usw.). Diese gruppierten Kosten- bzw. AU-Daten wurden als exakte Matching-Kriterien definiert (für einen PAIN2020-Interventionspatienten, der keine Kosten im Bereich der ambulanten Versorgung verursachte, wurde z.B. also ein Vergleichspatient gesucht, der in dieser Kostenkategorie ebenfalls keine Kosten verursacht hatte). Mit diesem Vorgehen konnte ein zufriedenstellendes Matchingmodell entwickelt werden. In **Tab. 4** sind die in diesem Matchingmodell verwendeten Merkmale und deren Verwendung als Prädiktoren im Propensity-Score-Matching bzw. zusätzlich als exakte Matchingkriterien aufgeführt.

Die im Folgenden gezeigten Ergebnisse wurden mit dem Analyse-Framework SAS Enterprise Guide 8.3 unter Verwendung der Prozedur PSMATCH mit der Methode „greedy“ und einer Ziehung von  $k = 1$  Zwillingen und einem Caliper = 0,2 erzeugt. Die ursprüngliche Stichprobe von 2,8 Mio. Versicherten wurde zur technischen Realisierung des Modells vor den weiteren Berechnungsschritten per Zufallsziehung auf 1,4 Mio. Versicherte reduziert.

Tab. 5: Standardisierte mittlere Differenzen zwischen Interventions- und gematchter Kontrollgruppe

	<b>Kodierung</b>	<b>Beobachtungen</b>	<b>Mittlere Differenz</b>	<b>Standardisierte Differenz</b>	<b>Prozent Reduktion</b>
Propensity Score		Alle	<0,0001	0,1463	
		Gematcht	<0,0001	-0,0004	99,71
<b>Soziodemographie</b>					
Alter	in Jahren	Alle	7,2546	0,5107	
		Gematcht	-0,8024	-0,0565	88,94
Geschlecht	0 = männlich	Alle	-0,0571	-0,1183	
	1 = weiblich	Gematcht	0,0000	0,0000	100,00
Familienversichert	0 = nein	Alle	0,0553	0,2036	
	1 = ja	Gematcht	0,0000	0,0000	100,00
<b>Schmerz</b>					
Kopfschmerz	0 = kein Auftreten	Alle	0,0500	0,1073	
	1 = Auftreten	Gematcht	-0,0102	-0,0219	79,63
Gelenkschmerz	0 = kein Auftreten	Alle	0,2256	0,4685	
	1 = Auftreten	Gematcht	-0,0102	-0,0212	95,49
Schmerzen Rücken/ Wirbelsäule	0 = kein Auftreten	Alle	0,2218	0,6173	
	1 = Auftreten	Gematcht	0,0102	0,0284	95,41
Schmerzen Weichteil- gewebe	0 = kein Auftreten	Alle	0,2267	0,4750	
	1 = Auftreten	Gematcht	-0,0122	-0,0256	94,61
Sonstige Schmerzen	0 = kein Auftreten	Alle	0,1621	0,4660	
	1 = Auftreten	Gematcht	0,0041	0,0117	97,49
<b>Leistungsbereich</b>					
AMB	in €	Alle	157,5693	0,5486	
		Gematcht	17,9131	0,0624	88,63
KHAMB	in €	Alle	25,1731	0,2862	
		Gematcht	-4,7869	-0,0544	80,98
KH	in €	Alle	50,5268	0,0832	
		Gematcht	-2,5214	-0,0042	95,01
AM	in €	Alle	11,1919	0,0873	
		Gematcht	-4,0092	-0,0313	64,18
HEILHILF	in €	Alle	100,1538	0,6705	
		Gematcht	10,4449	0,0699	89,57
REHA	in €	Alle	5,3201	0,0400	
		Gematcht	-4,3707	-0,0328	17,85
AU	in Tagen	Alle	3,0530	0,1964	
		Gematcht	0,0937	0,0060	96,93

AMB: Ambulante vertragsärztliche Versorgung, KHAMB: Ambulante Krankenhausleistungen, KH: Stationäre Krankenhausleistungen, AM: Arzneimittel, HEILHILF: Heil- und Hilfsmittel, REHA: Stationäre Rehabilitationsleistungen, AU: Arbeitsunfähigkeitstage.



Die Güteparameter des realisierten Matchingmodells sind in **Tab. 5** dargestellt. Es wurde für insgesamt  $n = 491$  Interventionsteilnehmer ein Matchingpartner gefunden. Gezeigt werden für alle im Matching-Verfahren berücksichtigten Merkmale die Unterschiede zwischen den Gruppen vor („Alle“) und nach dem Matching („Gematcht“) sowohl in natürlichen Einheiten als auch als standardisierte Differenzen. Es ist zu erkennen, dass das gewählte Matching-Verfahren zu einer substantiellen Verkleinerung der Unterschiede zwischen den Gruppen führte. Es ist davon auszugehen, dass die beiden Gruppen strukturell im Hinblick auf die verwendeten Kriterien hinreichend vergleichbar sind.

### zu 3) Vergleich der Durchschnittskosten von Interventionsgruppe und Matchingpartner in der Nachbeobachtung

Der letzte Arbeitsschritt bestand im Vergleich der durchschnittlichen Kosten und AU-Tage von Interventionsteilnehmern und gematchten Versicherten in der Nachbeobachtung. Für die Nachbeobachtung wurden die beiden Quartale berücksichtigt, die auf das Quartal des Assessments folgten.

## **3. Ergebnisse der Evaluation**

### **3.1 Evaluationsansatz 1**

#### **3.1.1 Fallzahlen**

Es beteiligten sich insgesamt 28 Zentren an PAIN2020. Im Zeitraum zwischen Februar 2019 und August 2021 wurden 1.233 Personen zu einem Informationstermin vor Ort in den Zentren eingeladen (**Abb. 3**).

223 (18,1%) der zu einem Informationstermin eingeladenen Personen nahmen wegen nicht erfüllter Einschlusskriterien oder wegen fehlender Einwilligung nicht an der Studie teil, sodass 1.010 Personen im Verhältnis von 70:30 entweder in die IG oder die KG randomisiert wurden.

Mit allen Studienteilnehmern wurde ein Assessment-Termin vereinbart (T1), in dessen Rahmen die primären Zielgrößen mittels des Deutschen Schmerzfragebogens erstmals erhoben wurden und für insgesamt 906 (89,7%) Personen vorlagen. Von dieser Ausgangsstichprobe ausgehend sind zur Nachbeobachtung nach drei Monaten (T2) 696 (76,8%) und zur Nachbeobachtung nach sechs Monaten (T3) 620 (68,4%) gültige Verlaufsfragebögen von den Studienteilnehmern zurückgesendet worden.

#### Ausschluss von der Datenanalyse

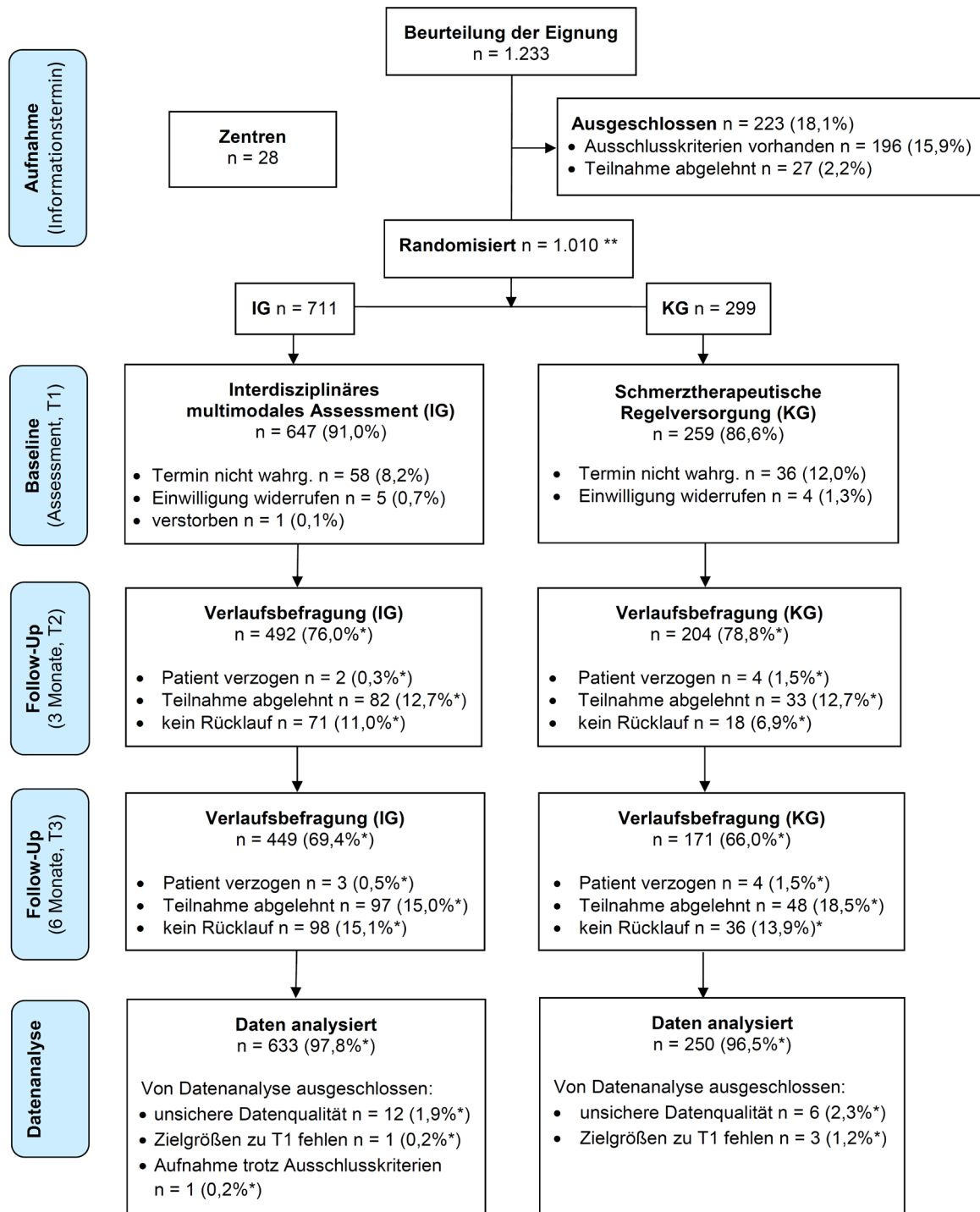
Aus Gründen der Qualitätssicherung wurden alle Fälle eines teilnehmenden Zentrums von der Datenanalyse ausgeschlossen (Zentrum J). Dies betrifft 18 (2,0%) Studienteilnehmer der Baseline-Stichprobe. Der Ausschluss von der Datenanalyse begründet sich im Detail durch:

- eine allgemein unsichere Datenqualität des Zentrums aufgrund systematischer Inkonsistenzen zwischen elektronisch und in Papierform vorliegenden Daten,
- nicht überprüfbare, da nur elektronisch vorliegende Zielgrößen zu T1,
- z.T. nicht vollzogene bzw. nicht termingerechte Aussendung von Verlaufsfragebögen,
- die Benutzung des für die Studie nicht vorgesehenen und nicht alle studienrelevanten Messinstrumente enthaltenden Routine-Verlaufsfragebogens der Deutschen Schmerzgesellschaft für die Nachbeobachtung zu T2 und T3.

Darüber hinaus wurden weitere 4 (0,4%) Fälle von der Datenanalyse ausgeschlossen, da bei ihnen zu T1 beide primäre Zielgrößen (SI, SBF) fehlten. Bei einem Fall (0,1%) erfolgte eine Studienaufnahme trotz bestehender Ausschlusskriterien. Dieser wurden ebenfalls nicht für

die Datenanalyse berücksichtigt. Somit standen insgesamt 883 Datensätze für die Auswertung zur Verfügung.

Abb. 3: Flussdiagramm zum Einschluss der PAIN2020-Teilnehmenden in die Evaluation (nach CONSORT)



IG: Interventionsgruppe, KG: Kontrollgruppe

\* bezogen auf T1

\*\* n = 4 Fälle sind fälschlicherweise randomisiert worden und nicht berücksichtigt



### 3.1.2 Drop-Outs

#### Studienabbrüche zwischen Aufnahme und Assessment (T1)

Nach erfolgter Studienaufnahme im Rahmen des Informationstermins erschienen 104 (10,3%) der bereits randomisierten Personen (zumeist aufgrund von nachträglichen Teilnahmeablehnungen bzw. eines expliziten Widerrufs der Einwilligung) nicht zum anschließenden Assessment-Termin als erstem Messzeitpunkt (**Abb. 3**). Die Wahrscheinlichkeit für einen solchen Studienabbruch fiel im Rahmen eines logistischen Regressionsmodells (dessen Erklärungsbeitrag insgesamt allerdings nur sehr gering war) in der KG mit einem Odds Ratio von 1,5 etwas höher aus als in der IG (**Tab. 6**).

Tab. 6: Vorhersage eines Studienabbruchs zwischen Aufnahme und Assessment (T1)

Prädiktor	Referenz <sup>a</sup>	OR	95%-CI	B	SE	p-Wert
(Konstante)		0,1		-2,0	0,206	<0,001
Gruppe	IG/KG	1,5	1,0 – 2,3	0,4	0,217	<0,05
Geschlecht	weiblich/männlich	0,8	0,5 – 1,2	-0,3	0,232	0,236
Alter	18-49 Jahre					0,357
	50-59 Jahre	0,7	0,4 – 1,2	-0,3	0,260	0,189
	über 60 Jahre	0,8	0,5 – 1,2	-0,3	0,251	0,253

n = 1.008, Pseudo-R<sup>2</sup> (Nagelkerke): 0,014 (p = 0,135) unter der Methode „Einschluss“, OR: Odds Ratio, CI: Konfidenzintervall, B: Regressionskoeffizient, SE: Standardfehler, IG: Interventionsgruppe, KG: Kontrollgruppe, <sup>a</sup> Referenzkategorie bei dichotomen Merkmalen ist jeweils die erstgenannte Merkmalsausprägung.

#### Rücklauf der Verlaufsbefragungen nach drei (T2) und sechs Monaten (T3)

Tab. 7: Vorhersage eines Drop-Outs zwischen Assessment (T1) und Nachbeobachtung nach 6 Monaten (T3)

Prädiktor	Referenz <sup>a</sup>	OR	95%-CI	B	SE	p-Wert
(Konstante)		1,3		0,3	0,329	0,421
Gruppe	IG/KG	1,2	0,8 – 1,6	0,1	0,173	0,414
Geschlecht	weiblich/männlich	1,2	0,8 – 1,6	0,1	0,164	0,372
Alter	18-49 Jahre					0,089
	50-59 Jahre	0,7	0,5 – 1,0	-0,4	0,194	0,057
	über 60 Jahre	0,7	0,5 – 1,0	-0,4	0,192	0,054
MCS (VR-12)		1,0	1,0 – 1,0	-0,0	0,007	0,003

n = 799, Pseudo-R<sup>2</sup> (Nagelkerke): 0,030 (p < 0,01), OR: Odds Ratio, CI: Konfidenzintervall, B: Regressionskoeffizient, SE: Standardfehler, IG: Interventionsgruppe, KG: Kontrollgruppe, <sup>a</sup> Referenzkategorie bei dichotomen Merkmalen ist jeweils die erstgenannte Merkmalsausprägung. Anm.: Die Variablen *Gruppe*, *Geschlecht* und *Alter* wurden als Kontrollvariablen per Einschluss in das Modell aufgenommen, während die zu T1 vorliegenden Instrumente *SBL*, *MFHW*, *DASS*, *PCS*, *VR-12*, *MPSS* sowie die *Schulbildung* per Rückwärtsselektion berücksichtigt wurden.

Bezogen auf alle für die Datenanalyse berücksichtigten 883 Fälle belief sich die durchschnittliche Quote zurückgesendeter Verlaufsfragebögen für T2 auf 77,8% und für T3 auf 69,4%<sup>3</sup>. Der tatsächliche Probandenverlust mit 22,2% zwischen T1 zu T2 sowie mit 10,8% zwischen T2 zu T3 fiel geringer aus, als er ursprünglich im Evaluationskonzept mit jeweils 30% bei jeder der beiden Verlaufsbefragungen erwartet wurde. Ein logistisches Regressionsmodell zur Vorhersage des Drop-Outs zwischen T1 und T3 identifizierte unter den zur Baseline eingesetzten Messinstrumenten sowie unter Berücksichtigung von Kontrollvariablen keine relevanten Einflussfaktoren (**Tab. 7**). In Bezug auf die einzelnen Zentren mit mindestens 10

<sup>3</sup> Von allen hier berichteten Rückläufen wurden 62 (9,0%) Verlaufsfragebögen zu T2 sowie 43 (7,0%) Verlaufsfragebögen zu T3 online ausgefüllt. Die Rücklaufquote für online ausgefüllte Verlaufsfragebögen unter allen per Email angeschriebenen Teilnehmenden betrug 69,8% zu T2 und 59,5% zu T3.

Studienteilnehmern zu T1 variierte die Rücklaufquote mit 54-96% zu T2 bzw. 27-96% zu T3 zum Teil stark (**Tab. 8**).

Tab. 8: Rücklauf der Verlaufsbefragungen nach 3 (T2) und 6 Monaten (T3) je Zentrum

Zentrum	Assessments (T1)	Verlauf nach 3 Monaten (T2)		Verlauf nach 6 Monaten (T3)	
	Fallzahl N	Rücklauf n	Rücklaufquote %	Rücklauf n	Rücklaufquote %
A	86	71	83	65	76
B	64	55	86	54	84
C	77	58	75	59	77
D	2	2		2	
E	23	21	91	22	96
F	30	28	93	26	87
G	22	17	77	12	55
H	74	46	62	29	39
I	21	17	81	17	81
K	26	14	54	7	27
L	29	17	59	13	45
M	7	5		0	
N	84	68	81	59	70
O	23	22	96	17	74
P	26	18	69	20	77
Q	46	33	72	39	85
R	59	42	71	41	69
S	35	25	71	24	69
T	12	9	75	8	67
U	39	37	95	32	82
V	27	25	93	21	78
W	3	2		0	
X	25	20	80	18	72
Y	11	9	82	6	55
Z	17	14	82	9	53
Ä	10	8	80	7	70
Ö	5	2		3	
<b>Gesamt</b>	<b>883</b>	<b>685</b>	<b>79</b>	<b>610</b>	<b>69</b>

Anm.: Die zentrumsbezogene Rücklaufquote ist für Zentren mit einer Baseline-Fallzahl von mindestens n = 10 Studienteilnehmenden ausgewiesen. Die Maxima und Minima einer Kategorie sind farblich grün bzw. rot hervorgehoben. Für die n = 18 zu T1 vorliegenden Fälle des von der Datenanalyse ausgeschlossenen Zentrums J waren n = 4 Verlaufsfragebögen zu T2 (Rücklauf 22,2%) und n = 3 Verlaufsfragebögen zu T3 (Rücklauf 16,7%) eingegangen.

### 3.1.3 Stichprobenmerkmale

Die Gesamtstichprobe zu T1 setzte sich zu zwei Dritteln aus Frauen zusammen (**Tab. 9**). Studienteilnehmer waren im Mittel 55,1 Jahre alt. Bei etwa der Hälfte betraf die Hauptschmerzdiagnose den Bereich der Wirbelsäule und des Rückens. Hinsichtlich der Zielgrößen wurden durch die Studienteilnehmer auf der Numerischen Ratingskala (0-10) eine im Durchschnitt mittelgradig ausgeprägte Schmerzintensität von 5,8 und eine etwas geringere schmerzbedingte Funktionseinschränkung von 4,9 berichtet. Die Merkmalsunterschiede zwischen Interventions- (IG) und Kontrollgruppe (KG) waren erwartungsgemäß gering und statistisch nicht bedeutsam.

Tab. 9: Stichprobenmerkmale

Merkmal		Interventions- gruppe (IG)		Kontroll- gruppe (KG)		p-Wert*
		n	%	n	%	
<b>Geschlecht</b>	<i>weiblich</i>	421	(66,5)	162	(64,8)	0,629
	<i>männlich</i>	212	(33,5)	88	(35,2)	
	<i>fehlend</i>	0	(0,0)	0	(0,0)	
<b>Alter</b>	MW ± SD (in Jahren)	55,0 ± 12,7		55,5 ± 13,4		0,589
	Min - Max (in Jahren)	20 - 87		22 - 89		
	<i>fehlend</i>	1	(0,2)	0	(0,0)	
<b>Schulbildung</b>	<i>kein Abschluss</i>	3	(0,5)	1	(0,4)	0,960
	<i>Hauptschule/Volksschule</i>	74	(11,8)	31	(12,6)	
	<i>Realschule/Mittlere Reife</i>	253	(40,2)	97	(39,3)	
	<i>Fachhochschulreife</i>	91	(14,5)	40	(16,2)	
	<i>Abitur/allg. Hochschulreife</i>	208	(33,1)	78	(31,6)	
	<i>fehlend</i>	4	(0,6)	3	(1,2)	
<b>Personen im Haushalt</b> (Mehrfachantworten möglich)	<i>ich lebe allein</i>	123	(20,2)	49	(19,6)	0,997
	<i>Ehepartner/Partner</i>	456	(74,9)	179	(71,6)	
	<i>Kinder</i>	174	(28,6)	65	(26,0)	
	<i>(Schwieger-)Eltern</i>	25	(4,1)	10	(4,0)	
	<i>fehlend</i>	24	(3,8)	12	(4,8)	
<b>berufstätig</b>	<i>ja</i>	427	(67,6)	180	(72,0)	0,200
	<i>nein</i>	205	(32,4)	70	(28,0)	
	<i>fehlend</i>	1	(0,2)	0	(0,0)	
<b>davon zurzeit arbeits- unfähig</b>	<i>ja</i>	43	(10,1)	16	(9,0)	0,664
	<i>nein</i>	381	(89,9)	162	(91,0)	
	<i>fehlend</i>	3	(0,7)	2	(1,1)	
<b>davon AU-Tage letzte 3 Monate</b>	MW ± SD (in Tagen)	10,9 ± 22,0		9,1 ± 20,4		0,372
	<i>fehlend</i>	13	(3,0)	10	(5,6)	
<b>Absicht, einen Renten- antrag zu stellen</b>	<i>ja</i>	22	(3,6)	7	(2,9)	0,630
	<i>nein</i>	595	(96,4)	234	(97,1)	
	<i>fehlend</i>	16	(2,5)	9	(3,6)	
<b>Bezug einer Rente</b>	<i>ja</i>	145	(27,0)	53	(24,9)	0,553
	<i>nein</i>	392	(73,0)	160	(75,1)	
	<i>fehlend</i>	96	(15,2)	37	(14,8)	
<b>Anerkannter Grad der Behinderung (GdB)</b>	<i>ja</i>	105	(17,7)	34	(14,3)	0,238
	<i>nein</i>	487	(82,3)	203	(85,7)	
	<i>fehlend</i>	41	(6,5)	13	(5,2)	
<b>Schmerzen seit...</b>	<i>weniger als 1 Monat</i>	2	(0,3)	0	(0,0)	0,554
	<i>1 Monat bis 1/2 Jahr</i>	77	(12,2)	39	(15,7)	
	<i>1/2 Jahr bis 1 Jahr</i>	180	(28,4)	72	(29,0)	
	<i>1 bis 2 Jahren</i>	160	(25,3)	52	(21,0)	
	<i>2 bis 5 Jahren</i>	105	(16,6)	43	(17,3)	
	<i>mehr als 5 Jahren</i>	109	(17,2)	42	(16,9)	
	<i>fehlend</i>	0	(0,0)	2	(0,8)	

Tab. 9: (Forts.)

Merkmal		Interventions- gruppe (IG)		Kontroll- gruppe (KG)		p-Wert*
		n	%	n	%	
		633	(100)	250	(100)	
<b>Hauptschmerz- diagnosen</b> (Mehrfachantworten möglich)	<i>Kopf-/Gesichtsschmerz</i>	50	(7,9)	10	(4,0)	0,161
	<i>Wirbelsäule/Rückenschmerz</i>	309	(48,8)	141	(56,4)	
	<i>Gelenkschmerz</i>	123	(19,4)	40	(16,0)	
	<i>Schmerz Muskel-Skelett-System/ Bindegewebschmerz</i>	109	(17,2)	42	(16,8)	
	<i>Ganzkörperschmerz/Fibromyalgie</i>	4	(0,6)	3	(1,2)	
	<i>Neuropathischer Schmerz</i>	21	(3,3)	5	(2,0)	
	<i>chronischer Schmerz ohne nähere Spezifikation</i>	16	(2,5)	10	(4,0)	
	<i>chron. Schmerzstörung mit soma- tischen und psych. Faktoren (F45.41)</i>	37	(5,8)	20	(8,0)	
	<i>psychische Verhaltensstörungen</i>	5	(0,8)	0	(0,0)	
	<i>Tumorschmerz</i>	1	(0,2)	1	(0,4)	
	<i>Brust-/Thoraxschmerz</i>	3	(0,5)	0	(0,0)	
	<i>Abdominalschmerz</i>	5	(0,8)	1	(0,4)	
	<i>Urogenitalschmerz</i>	2	(0,3)	1	(0,4)	
	<i>sonstige Schmerzen</i>	1	(0,2)	0	(0,0)	
		fehlend	5	(0,8)	3	
<b>Schmerzintensität</b>	MW ± SD (NRS, 0 - 10)	5,8 ± 1,7		5,9 ± 1,6		0,374
	fehlend	0	(0,0)	0	(0,0)	
<b>Schmerzbedingte Funktionseinschränkung</b>	MW ± SD (NRS, 0 - 10)	4,9 ± 2,3		4,8 ± 2,3		0,626
	fehlend	0	(0,0)	1	(0,4)	
<b>Schweregrad nach Von Korff</b>	0	3	(0,5)	0	(0,0)	0,234
	1	126	(20,1)	49	(19,9)	
	2	148	(23,6)	73	(29,7)	
	3	164	(26,1)	52	(21,1)	
	4	187	(29,8)	72	(29,3)	
	fehlend	5	(0,8)	4	(1,6)	
<b>Mainz Pain Staging System (MPSS)</b>	1	248	(39,3)	92	(37,4)	0,086
	2	317	(50,2)	115	(46,7)	
	3	66	(10,5)	39	(15,9)	
	fehlend	2	(0,3)	4	(1,6)	
<b>Schmerzbeschreibungs- liste (SBL)</b>	MW ± SD (affektive Summe, 0 - 12)	4,1 ± 3,4		4,1 ± 3,3		0,851
	fehlend	32	(5,1)	14	(5,6)	
<b>Habituelles Wohl- befinden (MFHW)</b>	MW ± SD (Summenwert, 0 - 35)	17,1 ± 8,3		17,0 ± 8,1		0,913
	fehlend	2	(0,3)	0	(0,0)	
<b>Depression Anxiety Stress Scale (DASS)</b>	MW ± SD (Depression, 0 - 21)	5,4 ± 4,5		5,5 ± 4,4		0,922
	MW ± SD (Angst, 0 - 21)	3,1 ± 3,2		2,9 ± 3,0		0,395
	MW ± SD (Stress, 0 - 21)	7,5 ± 4,6		7,9 ± 4,4		0,343
	fehlend	1	(0,2)	0	(0,0)	

Tab. 9: (Forts.)

Merkmal		Interventions- gruppe (IG)		Kontroll- gruppe (KG)		p-Wert*
		n	%	n	%	
<b>Pain Catastrophizing Scale</b>	MW ± SD (Hilflosigkeit (H), 0 - 20)	633	(100)	250	(100)	0,658
		7,9 ± 4,7		8,0 ± 4,5		
	fehlend	2	(0,3)	6	(2,4)	
	MW ± SD (Verstärkung (M), 0 - 12)	633	(100)	250	(100)	0,720
		4,1 ± 2,7		4,2 ± 2,6		
	fehlend	3	(0,5)	7	(2,8)	
	MW ± SD (Grübeln (R), 0 - 20)	633	(100)	250	(100)	0,874
		8,6 ± 4,9		8,5 ± 4,6		
	fehlend	5	(0,8)	8	(3,2)	
	MW ± SD (gesamt, 0 - 52)	633	(100)	250	(100)	0,859
	20,6 ± 11,0		20,7 ± 10,4			
fehlend	2	(0,3)	6	(2,4)		
<b>Gesundheitsbezogene Lebensqualität (VR-12)</b>	MW ± SD (PCS, Norm: 50)	633	(100)	250	(100)	0,498
		35,9 ± 9,1		35,4 ± 9,5		
	MW ± SD (MCS, Norm: 50)	633	(100)	250	(100)	0,958
	44,8 ± 11,5		44,8 ± 10,5			
fehlend	7	(1,1)	2	(0,8)		
<b>Medikamente vor Therapiebeginn</b>	<i>ja</i>	467	(75,5)	200	(81,3)	0,025
	<i>nein</i>	163	(24,5)	46	(18,7)	
	fehlend	3	(0,5)	4	(1,6)	

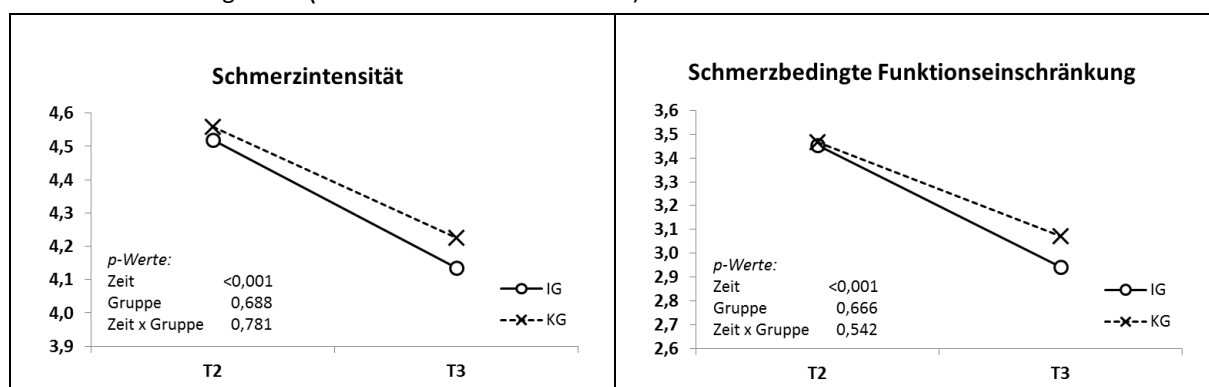
MW: Mittelwert, SD: Standardabweichung, Min: Minimum, Max: Maximum, NRS: Numerische Ratingskala, AU: Arbeitsunfähigkeit. \* Trotz randomisierter kontrollierter Studie erfolgten informationshalber Tests auf Unterschiede zwischen IG und KG mittels t-Test für intervallskalierte Merkmale und Chi-Quadrat-Test für kategoriale Merkmale zum Signifikanzniveau 0,05. Anm.: Prozentangaben bei „fehlend“ beziehen sich auf die gesamte Gruppenstichprobe; kursive Prozentwerte geben die gültigen Prozentwerte der Verteilung des jeweiligen Merkmals wieder.

### 3.1.4 Veränderung der primären Zielgrößen im Zeitverlauf

#### Auswertungsvariante I: Zufallskoeffizientenmodelle

Die folgenden Abbildungen zeigen die unter dem weiter oben bereits beschriebenen Modell der Kovarianzanalyse vorhergesagten Werte der primären Zielgrößen (**Abb. 4**). Zur Information über die statistische Bedeutsamkeit der untersuchten Effekte werden in den Abbildungen nur die p-Werte für die primär interessierenden Effekte (Zeit, Gruppenzugehörigkeit und deren Interaktion) dargestellt. Die vollständigen Ergebnisse werden im Anhang (**Anhang 3**) wieder-gegeben.

Abb. 4: Primäre Zielgrößen (I Zufallskoeffizientenmodelle)



Wie sich zeigt, wurden für den Messzeitpunkt T2 für beide Gruppen nahezu identische Werte der primären Zielgrößen vorhergesagt. Für den Messzeitpunkt T3 wurden nur unwesentlich niedrigere Werte in der Interventionsgruppe erwartet. Bei beiden primären Zielgrößen waren statistisch signifikante Unterschiede in den Werteausprägungen an den Zeitpunkten T2 und

T3 zu beobachten. Die Unterschiede zwischen den Gruppen und der Interaktionseffekt erreichten jedoch bei beiden primären Zielgrößen keine statistische Signifikanz.

### Auswertungsvariante II: Complete-Case-Analyse

Im Rahmen des zweiten Auswertungsansatzes zeigten sich die Reduktionen der SI bzw. SBF nach drei bzw. sechs Monaten deskriptiv in ähnlichen Größenordnungen wie in der ersten Auswertungsvariante (**Tab. 10**).

Tab. 10: Veränderung der primären Zielgrößen im Zeitverlauf (II Complete Cases)

<b>Schmerzintensität (NRS, 0-10)</b>						
Messzeitpunkt	Interventionsgruppe (IG)		Kontrollgruppe (KG)		p-Wert t-test	Cohens d
	MW ± SD	n	MW ± SD	n		
<b>T1</b>	5,7 ± 1,7	(395)	5,9 ± 1,7	(153)	0,177	-0,129
<b>T2</b>	4,5 ± 2,2	(395)	4,5 ± 2,1	(153)	0,681	-0,039
<b>T3</b>	4,1 ± 2,3	(395)	4,2 ± 2,4	(153)	0,619	-0,047
<b>Δ<sub>T1-T2</sub></b>	-1,3 ± 2,3	(395)	-1,4 ± 2,1	(153)	0,535	0,059
<b>Δ<sub>T1-T3</sub></b>	-1,6 ± 2,4	(395)	-1,8 ± 2,5	(153)	0,645	0,044

<b>Schmerzbedingte Funktionseinschränkung (NRS, 0-10)</b>						
Messzeitpunkt	Interventionsgruppe (IG)		Kontrollgruppe (KG)		p-Wert t-test	Cohens d
	MW ± SD	n	MW ± SD	n		
<b>T1</b>	4,9 ± 2,4	(393)	4,8 ± 2,3	(155)	0,656	0,042
<b>T2</b>	3,4 ± 2,4	(393)	3,4 ± 2,4	(155)	0,686	0,038
<b>T3</b>	2,9 ± 2,3	(393)	2,9 ± 2,5	(155)	0,951	-0,006
<b>Δ<sub>T1-T2</sub></b>	-1,4 ± 2,6	(393)	-1,4 ± 2,5	(155)	0,982	-0,002
<b>Δ<sub>T1-T3</sub></b>	-2,0 ± 2,7	(393)	-1,8 ± 2,6	(155)	0,653	-0,043

MW: Mittelwert, SD: Standardabweichung, NRS: Numerische Ratingskala.

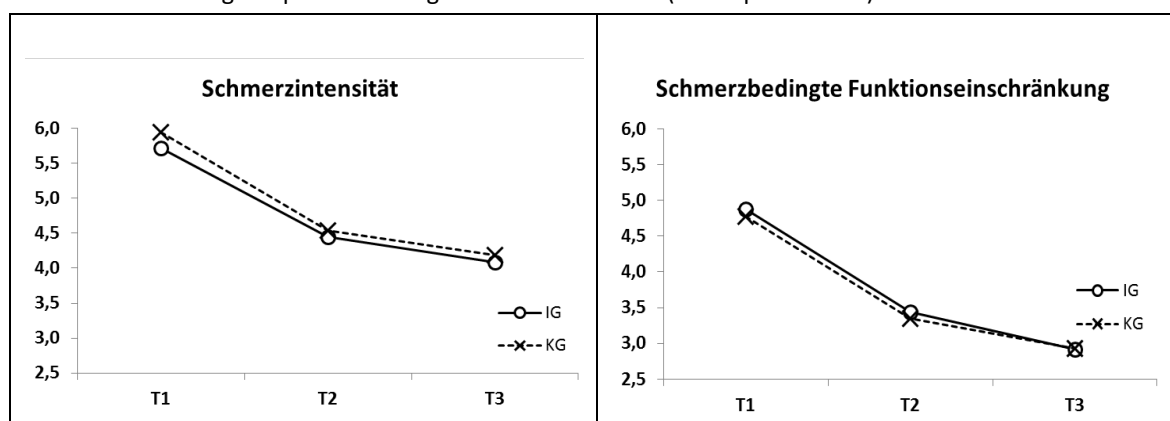
Tab. 11: Varianzanalyse mit Messwiederholung für die primären Zielgrößen (II Complete Cases)

Interaktionseffekt des Gruppenmerkmals mit...	n <sub>IG</sub>	n <sub>KG</sub>	F-Wert	p-Wert	Cohens f
<b>Schmerzintensität</b>	395	153	(1,91; 1.042,25) = 0,22	0,794	0,020
<b>Schmerzbedingte Funktionseinschränkung</b>	393	155	(1,87; 1.018,05) = 0,15	0,845	0,017

IG: Interventionsgruppe, KG: Kontrollgruppe.

Die (Greenhouse-Geisser-korrigierten) Varianzanalysen mit Messwiederholung zeigten weder für die SI noch die SBF einen statistisch signifikanten Interaktionseffekt mit dem Gruppenmerkmal an. Die gemessenen Effekte (Cohens f) sind entsprechend als sehr gering einzuschätzen (**Tab. 11**). Im Verlaufsdiagramm äußerten sich diese Befunde in jeweils sehr nahe beieinander verlaufenden Linien (**Abb. 5**).

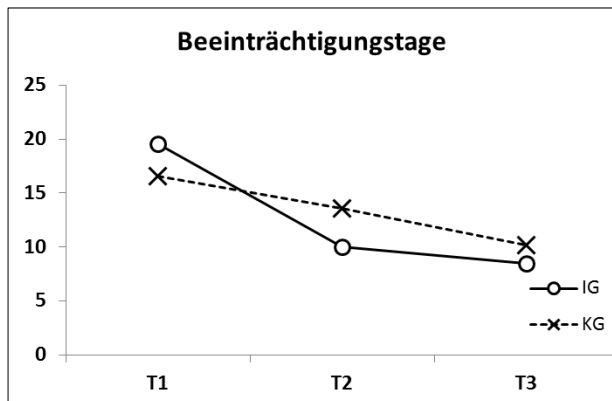
Abb. 5: Veränderung der primären Zielgrößen im Zeitverlauf (II Complete Cases)



IG: Interventionsgruppe, KG: Kontrollgruppe.

Ergänzend wurde eine Varianzanalyse mit Messwiederholung hinsichtlich der *Beeinträchtigungstage* durch Schmerzen (Item des Von-Korff-Fragebogens, das jedoch nicht in die Berechnung der Zielgrößen SI und SBF einfließt), durchgeführt. Für den Interaktionseffekt mit der Gruppenzugehörigkeit zeigte sich eine knappe statistische Signifikanz zum Signifikanzniveau  $\alpha = 0,05$  mit geringer Effektstärke ( $F(1,46; 718,75) = 3,59$ ;  $p = 0,042$ ; Cohens  $f = 0,085$ ). Absolut stand nach sechs Monaten einer Reduktion um durchschnittlich  $12,0 \pm 30,9$  Beeinträchtigungstage in der IG ( $n_{IG} = 406$ ) eine Verringerung von  $6,9 \pm 27,8$  Beeinträchtigungstagen in der KG gegenüber ( $n_{KG} = 159$ ) (**Abb. 6**). Dieser Effekt äußerte sich allerdings in einem insgesamt als gering zu bewertenden Cohens  $f < 0,1$ .

Abb. 6: Veränderung der Beeinträchtigungstage im Zeitverlauf (II Complete Cases)



IG: Interventionsgruppe, KG: Kontrollgruppe.

#### Auswertungsvariante III: Available Cases

Für die primäre Zielgröße SI konnten nach drei Monaten sowohl in der IG als auch in der KG durchschnittliche Reduktionen um 1,3 Punkte (ca. 22%) auf der Numerischen Ratingskala (NRS, 0-10) beobachtet werden (**Tab. 10**). Im gesamten sechsmonatigen Nachbeobachtungszeitraum reduzierte sich die SI in beiden Gruppen in derselben Größenordnung um 1,6 bzw. 1,7 Punkte (ca. 28% bzw. 29%). Die mittlere Verringerung der SI nach sechs Monaten entsprach *innerhalb der IG* einer Effektstärke (Cohens  $d$ ) von  $d = 0,661$  ( $p < 0,001$ ) und *innerhalb der KG* von  $d = 0,685$  ( $p < 0,001$ ).

Bei direktem Vergleich der mittleren Reduktionen der SI beider Gruppen miteinander (Vergleich der Mittelwertdifferenzen) bestand bei Effektstärken von jeweils  $|d| < 0,1$  zu keinem der Nachbeobachtungszeitpunkte ein statistisch bedeutsamer Unterschied (**Tab. 12**).

Ein ähnliches Bild zeigte sich auch für die zweite primäre Zielgröße SBF, welche sich nach drei Monaten in beiden Gruppen um 1,4 Punkte (ca. 29%) und nach sechs Monaten um 1,9 bzw. 1,8 Punkte (39% bzw. 38%) verringerte. *Innerhalb der Gruppen* bemaß sich die Effektstärke nach sechs Monaten auf  $d = 0,720$  ( $p < 0,001$ ) in der IG und auf  $d = 0,698$  ( $p < 0,001$ ) in der KG. Der Vergleich der Mittelwertdifferenzen miteinander zeigte auch für die SBF keine statistisch relevanten Verringerungen zugunsten einer der beiden Gruppen sowie ebenfalls nur marginale Effektstärken von jeweils  $|d| < 0,10$ .



Tab. 12: Veränderung der primären Zielgrößen im Zeitverlauf (III Available Cases)

Schmerzintensität (NRS, 0-10)						
Messzeitpunkt	Interventionsgruppe (IG)		Kontrollgruppe (KG)		p-Wert t-test	Cohens d
	MW ± SD	n	MW ± SD	n		
T1	5,8 ± 1,7	(633)	5,9 ± 1,6	(250)	0,374	-0,066
T2	4,5 ± 2,2	(482)	4,6 ± 2,2	(193)	0,679	-0,035
T3	4,1 ± 2,3	(438)	4,3 ± 2,4	(167)	0,419	-0,074
$\Delta_{T1-T2}$	-1,3 ± 2,3	(482)	-1,3 ± 2,2	(193)	0,673	0,036
$\Delta_{T1-T3}$	-1,6 ± 2,5	(438)	-1,7 ± 2,4	(167)	0,908	0,010

Schmerzbedingte Funktionseinschränkung (NRS, 0-10)						
Messzeitpunkt	Interventionsgruppe (IG)		Kontrollgruppe (KG)		p-Wert t-test	Cohens d
	MW ± SD	n	MW ± SD	n		
T1	4,9 ± 2,3	(633)	4,8 ± 2,3	(249)	0,626	0,036
T2	3,5 ± 2,5	(482)	3,4 ± 2,5	(196)	0,736	0,029
T3	2,9 ± 2,4	(436)	3,1 ± 2,6	(167)	0,598	-0,048
$\Delta_{T1-T2}$	-1,4 ± 2,5	(482)	-1,4 ± 2,5	(195)	0,740	-0,028
$\Delta_{T1-T3}$	-1,9 ± 2,7	(436)	-1,8 ± 2,6	(166)	0,560	-0,053

MW: Mittelwert, SD: Standardabweichung, NRS: Numerische Ratingskala.

### 3.1.5 Ergänzende Analysen

Da sich im Rahmen der vorangegangenen Analysen bzgl. der primären Zielgrößen keine relevanten Unterschiede zwischen der IG und der KG zeigten, werden im Folgenden einige ergänzende Analysen als potenzielle Erklärungsansätze für diese Beobachtungen dargestellt.

#### Herausnahme von Studienteilnehmern ohne Beschwerden nach 6 Monaten

Der drei bzw. sechs Monate nach dem Assessment für die primäre Datenerhebung in dieser Studie (aber auch in der Routineversorgung) eingesetzte Schmerz-Verlaufsfragebogen der Deutschen Schmerzgesellschaft e.V. (VFB) eruiert zunächst mit einer Eingangsfrage, ob die behandelten Schmerzen noch bestehen. Zwischen IG und KG bestand zu beiden Nachbeobachtungsmesszeitpunkten diesbezüglich kein statistisch relevanter Unterschied im Anteil jener Studienteilnehmer, die von keinen Schmerzen mehr berichteten (**Tab. 13**).

Tab. 13: Anteil der Studienteilnehmer mit und ohne Schmerzen bei Nachbeobachtung

Messzeitpunkt		IG		KG		p-Wert Chi <sup>2</sup> -Test
		n	%	n	%	
T1		633		250		
T2	keine Schmerzen mehr	47	(9,7)	18	(9,1)	0,821
	noch Schmerzen	440	(90,3)	180	(90,9)	
T3	keine Schmerzen mehr	60	(13,6)	26	(15,5)	0,547
	noch Schmerzen	382	(86,4)	142	(84,5)	

IG: Interventionsgruppe, KG: Kontrollgruppe.

Der VFB sieht an dieser Stelle eine Sprunganweisung vor, sodass Studienteilnehmer, die von keinen Schmerzen mehr berichteten, regulär keine Angaben u.a. zur Schmerzintensität und schmerzbedingten Funktionseinschränkung machen. Bei allen voranstehenden Analysen wurde mithin für diese Gruppe von Teilnehmenden jeweils ein Wert von „0“ für die SI sowie SBF postuliert.



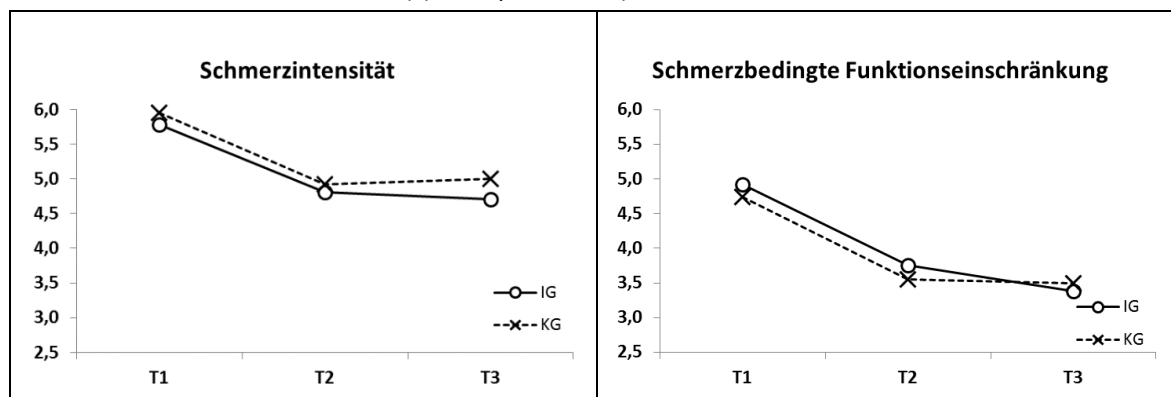
Tab. 14: Varianzanalyse mit Messwiederholung für die primären Zielgrößen (für Studienteilnehmer mit bestehenden Schmerzen nach 6 Monaten) (II Complete Cases)

Interaktionseffekt des Gruppenmerkmals mit...	n <sub>IG</sub>	n <sub>KG</sub>	F-Wert	p-Wert	Cohens f
Schmerzintensität	342	128	(1,96; 916,33) = 0,49	0,607	0,032
Schmerzbedingte Funktionseinschränkung	340	130	(1,89; 882,97) = 1,21	0,298	0,051

IG: Interventionsgruppe, KG: Kontrollgruppe.

Werden nur Studienteilnehmer mit bestehenden Schmerzen nach 6 Monaten in die Varianzanalyse mit Messwiederholung eingeschlossen, ergaben sich zwar weiterhin keine signifikanten Interaktionseffekte mit der Gruppenzugehörigkeit, jedoch etwas höhere Effektstärken, insbesondere für die SBF (**Tab. 14**). Im Verlaufsdigramm bewegten sich die Zielgrößen für beide Gruppen bis zum zweiten Messzeitpunkt nach drei Monaten zunächst im Gleichschritt, ehe sie zum dritten Messzeitpunkt nach sechs Monaten tendenziell zugunsten der IG auseinander zu divergieren scheinen (**Abb. 7**).

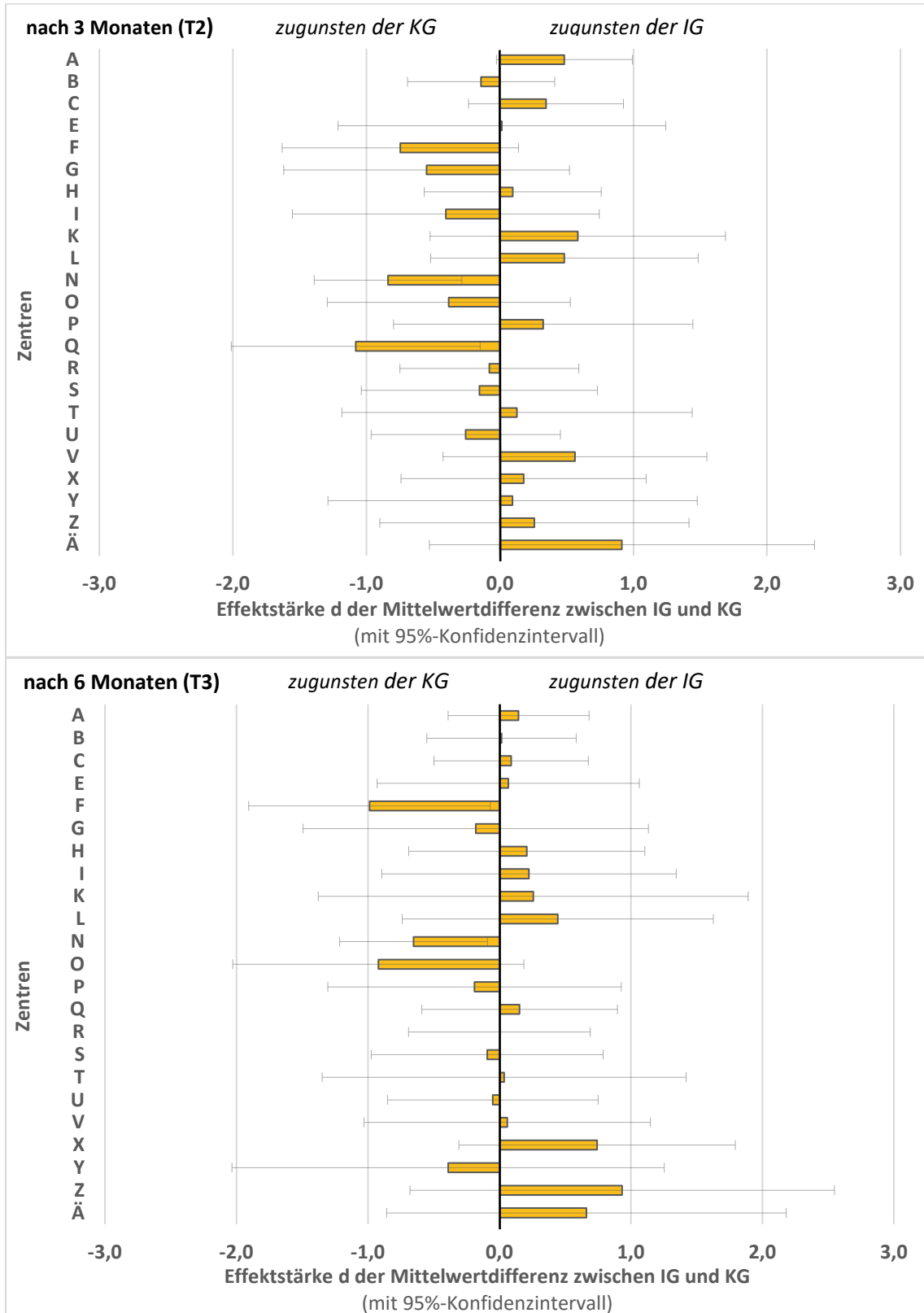
Abb. 7: Veränderung der primären Zielgrößen im Zeitverlauf (für Studienteilnehmer mit bestehenden Schmerzen nach 6 Monaten) (II Complete Cases)



IG: Interventionsgruppe, KG: Kontrollgruppe.

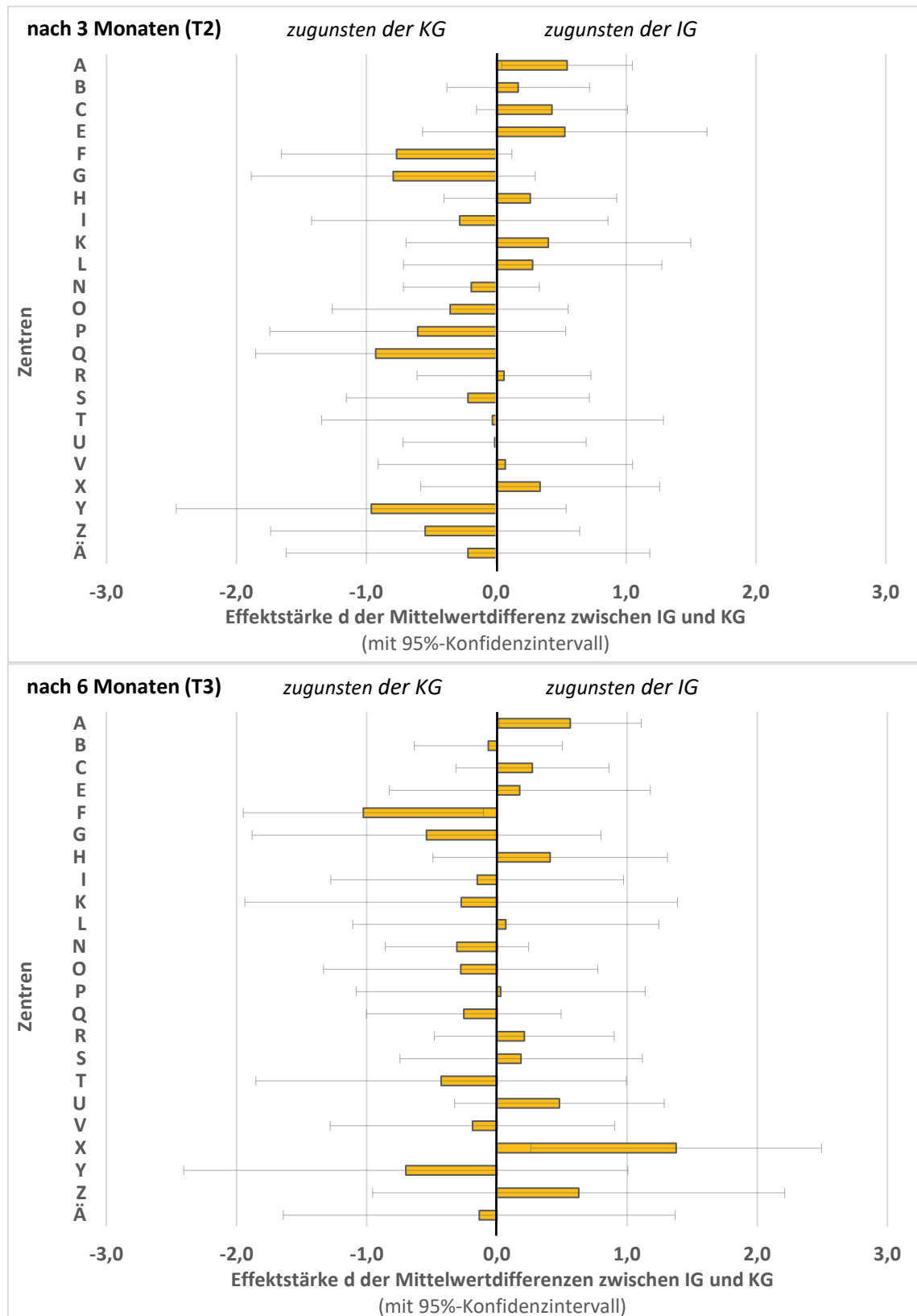
Zentrumsspezifische Auswertungen

Abb. 8: Zentrumsbezogene Effektstärken bzgl. der Schmerzintensität nach 3 und 6 Monaten (III Available Cases)



IG: Interventionsgruppe, KG: Kontrollgruppe. Mangels Fallzahlen fehlen die Zentren D, M, W und Ö.

Abb. 9: Zentrumsbezogene Effektstärken bzgl. der Schmerzbed. Funktionseinschränkung nach 3 und 6 Monaten (III Available Cases)



IG: Interventionsgruppe, KG: Kontrollgruppe. Mangels Fallzahlen fehlen die Zentren D, M, W und Ö.

Für die **Abb. 8** und **9** sind jeweils die Effektstärken der Mittelwertdifferenzen zwischen IG und KG für die beiden Zielgrößen und Nachbeobachtungszeiträume in Abhängigkeit der Zentren berechnet worden. Alle verfügbaren Fälle wurden berücksichtigt, es sei denn für ein Zentrum lagen zu einem Messzeitpunkt innerhalb einer Gruppe keinerlei Fälle vor. Aufgrund der meist sehr geringen Fallzahlen pro Zentrum (vgl. **Tab. 6**) und Gruppe fielen die 95%-Konfidenzintervalle zum Teil sehr breit aus. Ungeachtet dessen ergab sich ein sehr heterogenes Bild der verschiedenen Zentren, bei dem Effektstärken teils bis über 1,0 zugunsten der IG wie auch zugunsten der KG ermittelt wurden.

Vor diesem Hintergrund gab es folgende Überlegungen, die Zentren nach objektiven Strukturmerkmalen jeweils zu dichotomisieren:

- Zentren mit mindestens 30 versus Zentren mit weniger durchgeführten Assessments,
- Zentren mit ambulanter versus stationärer/teilstationärer Versorgung,
- Zentren mit hausinterner versus externer Durchführung der schmerztherapeutischen Regelversorgung,
- im Falle hausinterner schmerztherapeutischer Regelversorgung: Zentren, bei denen das Assessment in der IG und der KG ärztlicherseits von derselben Person durchgeführt wurde.

Bei keiner dieser Gruppenegegenüberstellungen ergab sich jedoch eine Systematik, welche die Heterogenität der berichteten zentrumsbezogenen Effektstärken stringent erklären konnte.

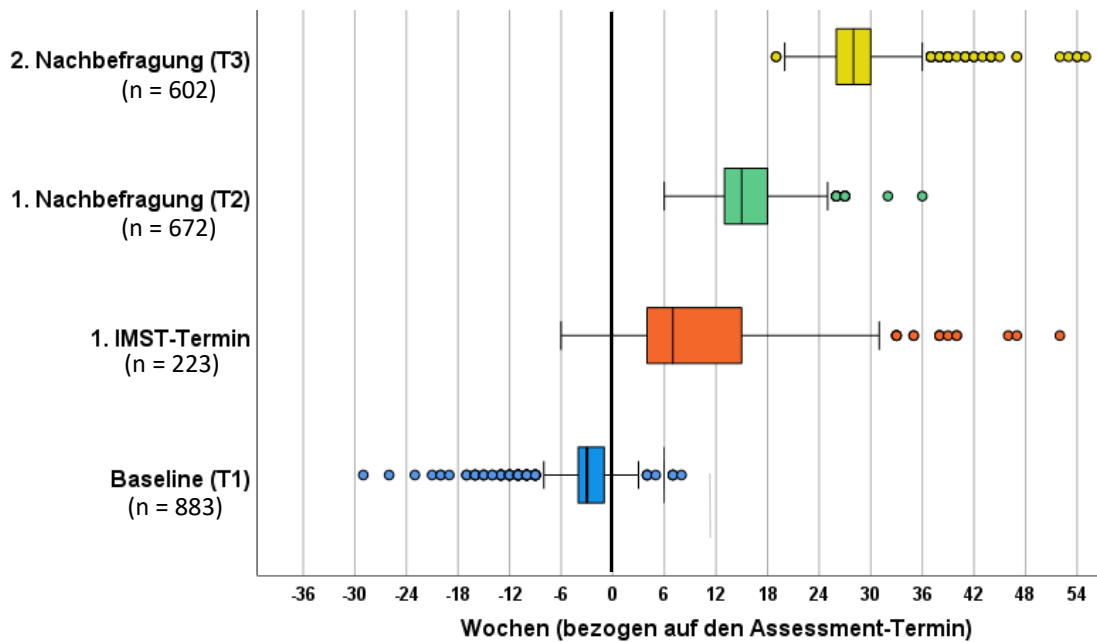
### Protokolladhärenz

Ausgehend vom Assessment-Termin als Indexereignis wurden in **Abb. 10** die Zeitabstände zu den in der Datenbank jeweils dokumentierten verschiedenen Erhebungszeitpunkten T1, T2 und T3 dargestellt. Hinsichtlich der Behandlung bot die Primärdatenbank zusätzlich die Terminangaben der für PAIN2020 konzipierten edukativen bzw. begleitenden interdisziplinären multimodalen Schmerztherapie (IMST). Ausreißer lassen vermuten, dass sich auf Einzelfallebene unterschiedliche Konstellationen ergaben, in denen die Datenerhebung zeitlich nicht dem Studienprotokoll entsprach.

Vor diesem Hintergrund wurden folgende fünf Fallkonstellationen als „protokoll-inadhärent“ definiert:

- A: Die Datenerhebung für den Ausgangswert (T1) erfolgte mehr als 8 Wochen vor dem Assessment-Termin.
- B: Die 2. Nachbefragung (T3) erfolgte mehr als 34 Wochen nach dem Assessment.
- C: Der 1. IMST-Termin erfolgte 12 Wochen oder später nach dem Assessment.
- D: Zwischen 1. IMST-Termin und 2. Nachbefragung (T3) lagen weniger als 12 Wochen.
- E: Für Studienteilnehmer der Interventionsgruppe, die eine Empfehlung zur edukativen oder begleitenden IMST erhalten haben, war kein entsprechender Termin dokumentiert.

Abb. 10: Boxplots dokumentierter Datenerhebungszeitpunkte in Abhängigkeit vom Assessment-Termin



Anmerkung: „1. IMST-Termin“ meint den ersten Termin, der auf Einzelfallebene im Rahmen einer edukativen oder begleitenden, interdisziplinären multimodalen Schmerztherapie dokumentiert war.

Es zeigte sich, dass etwas höhere (jedoch statistisch nicht signifikante) Effektstärken erreicht wurden, wenn insbesondere die Fälle aus Konstellation C von der Datenanalyse ausgeschlossen wurden (**Tab. 15**).

Tab. 15: Effektstärke für die primären Zielgrößen nach 6 Monaten bei Ausschluss protokoll-inadhärenter Fallkonstellationen (III Available Cases)

Schmerzintensität (NRS, 0-10)			
Ausgeschlossene Konstellation	Verbleibende Fälle n	Cohens d*	p-Wert
keine	605	0,010	0,908
A	561	0,042	0,657
B	550	-0,019	0,841
C	548	-0,053	0,571
D	562	-0,028	0,758
E	525	0,013	0,890

Schmerzbedingte Funktionseinschränkung (NRS, 0-10)			
Ausgeschlossene Konstellation	Verbleibende Fälle n	Cohens d*	p-Wert
keine	602	-0,053	0,560
A	558	-0,024	0,803
B	547	-0,099	0,300
C	546	-0,106	0,257
D	560	-0,093	0,317
E	523	-0,049	0,603

\* für die Mittelwertdifferenzen zwischen Interventions- und Kontrollgruppe

### 3.1.6 Veränderung der sekundären Zielgrößen im Zeitverlauf

#### Auswertungsvariante I: Zufallskoeffizientenmodelle

Die folgenden Abbildungen zeigen die unter dem weiter oben beschriebenen Modell vorhergesagten Werte der sekundären Zielgrößen. Wie bereits bei der Darstellung der Ergebnisse für die primären Zielgrößen werden in den Abbildungen wiederum nur die p-Werte für die primär interessierenden Effekte (Zeit, Gruppenzugehörigkeit und deren Interaktion) dargestellt. Die vollständigen Ergebnisse können dem Anhang (**Anhang 3**) entnommen werden.

Abb. 11a: Sekundäre Zielgrößen (I Zufallskoeffizientenmodelle)\*)

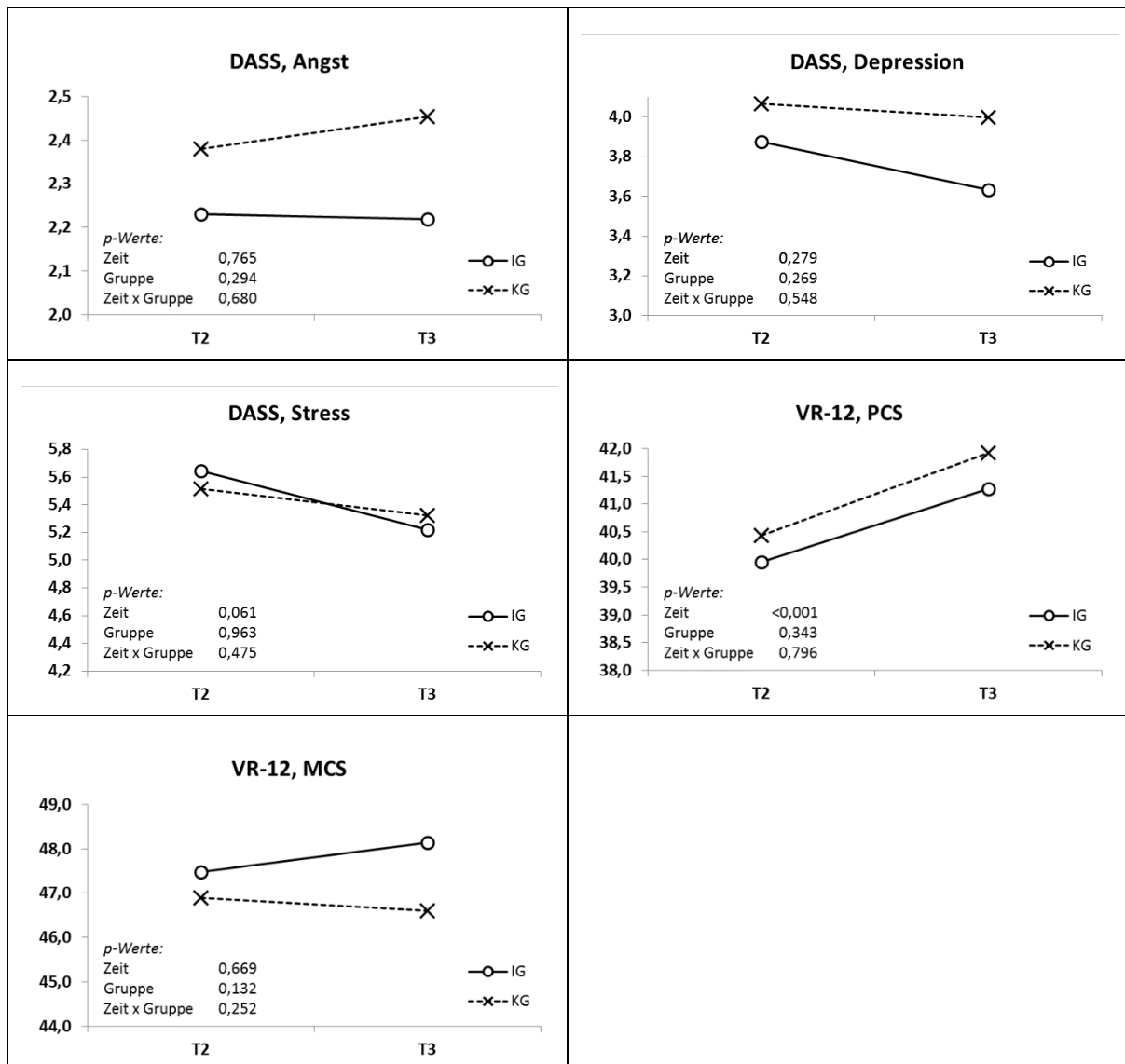
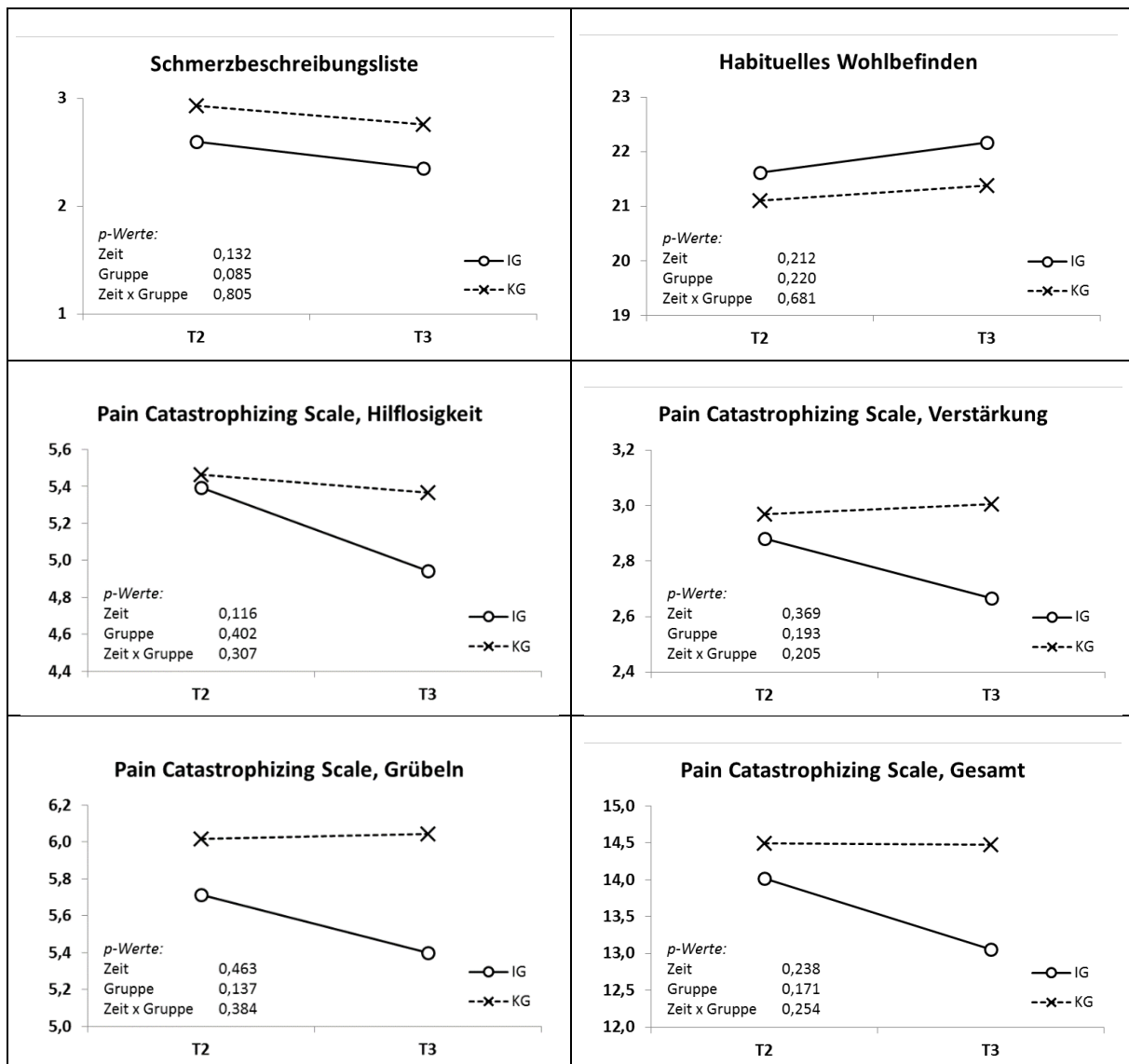


Abb. 11b: Zusätzlich ausgewertete sekundäre Zielgrößen (I Zufallskoeffizientenmodelle)\*)



\*) Bei der Interpretation der Ergebnisse ist zu beachten, dass bei den meisten Skalen hohe Werte eine ungünstige Ausprägung des Merkmals repräsentieren. In diesen Fällen impliziert eine Verlaufskurve für die Interventionsgruppe, die unterhalb der Kurve der Kontrollgruppe liegt, einen deskriptiven Unterschied zugunsten der Interventionsgruppe. Bei den folgenden Merkmalen entsprechen hohe Werte jedoch günstigen Ausprägungen, sodass hier genau der umgekehrte Fall vorliegt: Habituelleres Wohlbefinden, VR-12 PCS (Physical Component Summary), VR-12 MCS (Mental Component Summary).

Die Ergebnisse für die sekundären Zielgrößen zeigten bis auf wenige Ausnahmen (u.a. VR-12 PCS) zumindest deskriptiv erwartungskonforme Unterschiede zwischen den Gruppen zugunsten der IG. Diese Unterschiede waren bei den vorhergesagten Werten zu T3 in einigen Fällen deutlicher ausgeprägt als zu T2.

Nach den inferenzstatistischen Ergebnissen für die sekundären Zielgrößen konnten jedoch auch hier keine Unterschiede statistisch abgesichert werden, obwohl sich einige der p-Werte der Grenze von 0,05 näherten (z.B. Schmerzbeschreibungsliste SBL).

Betrachtet man über diese Ergebnisse hinaus die praktische Bedeutsamkeit der deskriptiven Unterschiede zwischen den Gruppen, so war in den meisten Fällen festzustellen, dass die Effektgrößen (berechnet anhand der Differenzen der erwarteten Gruppenmittelwerte und der gepoolten Standardabweichung des ersten Messzeitpunkts) in einem niedrigen Bereich von unter 0,10 lagen. Bei einigen Skalen, insbesondere bei den psychologischen Konstrukten (u.a.

Pain Catastrophizing-Skalen, VR-12 MCS) sowie der Schmerzbeschreibungsliste SBL, traten jedoch auch Effektgrößen geringfügig über 0,10 auf.

### Auswertungsvariante II: Complete-Case-Analyse

Die Überprüfung des Interaktionseffektes der Gruppenzugehörigkeit mit den jeweils eingesetzten Messinstrumenten mittels Varianzanalysen mit Messwiederholung (Greenhouse-Geisser-korrigiert) zeigte keine statistisch bedeutsamen Zusammenhänge (**Tab. 16**). Die berechneten Effektstärken lagen jeweils unterhalb  $f < 0,10$  als Schwelle für geringe Effekte.

Tab. 16a: Varianzanalyse mit Messwiederholung für die sekundären Zielgrößen (II Complete Cases)

Interaktionseffekt des Gruppenmerkmals mit...	n <sub>IG</sub>	n <sub>KG</sub>	F-Wert	p-Wert	Cohens f
DASS, Angst	395	153	(1,89; 1.033,32) = 2,12	0,123	0,063
DASS, Depression	395	153	(1,88; 1.026,05) = 0,54	0,572	0,032
DASS, Stress	394	153	(1,95; 1.061,89) = 0,06	0,963	0,011
VR-12, PCS	386	152	(1,89; 1.015,24) = 0,37	0,680	0,026
VR-12, MCS	386	152	(1,90; 1.020,02) = 0,89	0,408	0,042

IG: Interventionsgruppe, KG: Kontrollgruppe, PCS (VR-12): Physical Component Summary, MCS (VR-12): Mental Component Summary, DASS: Depression Anxiety Stress Scale.

Tab. 16b: Varianzanalyse mit Messwiederholung für die zusätzlich ausgewerteten sekundären Zielgrößen (II Complete Cases)

Interaktionseffekt des Gruppenmerkmals mit...	n <sub>IG</sub>	n <sub>KG</sub>	F-Wert	p-Wert	Cohens f
Schmerzbeschreibungsliste (SBL)	317	129	(1,89; 836,99) = < 0,01	0,998	0,002
Habituelles Wohlbefinden (MFHW)	393	153	(1,89; 1.026,71) = 0,86	0,428	0,045
Pain Catastrophizing Scale, Hilflosigk.	386	147	(1,90; 1.010,55) = 0,48	0,610	0,030
Pain Catastrophizing Scale, Verstärk.	383	145	(1,88; 990,80) = 1,16	0,313	0,047
Pain Catastrophizing Scale, Grübeln	374	143	(1,91; 984,31) = 0,50	0,594	0,031
Pain Catastrophizing Scale, Gesamt	384	147	(1,92; 1.015,83) = 0,78	0,456	0,038

IG: Interventionsgruppe, KG: Kontrollgruppe.

### Auswertungsvariante III: Available-Case-Analyse

Nach sechs Monaten zeigten sich alle sekundären Zielgrößen im Zeitverlauf verbessert (**Tab. 17**). Diese Verbesserungen hatten sich zu überwiegendem Teil jeweils in den ersten drei Monaten manifestiert. Im direkten Vergleich der Mittelwertdifferenzen zwischen IG und KG waren jedoch für die eingesetzten Messinstrumente keine statistisch relevanten Unterschiede zu beobachten, wenngleich sich die IG, absolut gesehen, nach sechs Monaten in fast allen Fällen etwas stärker verbesserte als die KG. Lediglich für die körperliche Summenskala des VR-12 trat nach sechs Monaten ein etwas stärkerer ausgeprägter Effekt zugunsten der KG auf.



Tab. 17a: Veränderung der sekundären Zielgrößen nach 3 und 6 Monaten (III Available Cases)

Instrument	Messung	IG		KG		p-Wert t-test	Cohens d
		MW ± SD	n	MW ± SD	n		
<b>DASS, Angst</b> (0-21)	$\Delta_{T1-T2}$	0,7 ± 2,8	(482)	0,4 ± 2,7	(193)	0,225	0,103
	$\Delta_{T1-T3}$	0,7 ± 2,8	(436)	0,3 ± 3,0	(166)	0,092	0,154
<b>DASS, Depression</b> (0-21)	$\Delta_{T1-T2}$	1,4 ± 3,7	(482)	1,2 ± 3,8	(193)	0,581	0,047
	$\Delta_{T1-T3}$	1,6 ± 3,9	(436)	1,3 ± 4,1	(166)	0,284	0,098
<b>DASS, Stress</b> (0-21)	$\Delta_{T1-T2}$	1,9 ± 4,0	(481)	2,1 ± 4,1	(193)	0,560	-0,050
	$\Delta_{T1-T3}$	2,3 ± 3,9	(436)	2,3 ± 4,5	(166)	0,991	-0,001
<b>VR-12, PCS</b> (Norm: 50)	$\Delta_{T1-T2}$	4,1 ± 8,5	(475)	4,7 ± 8,6	(192)	0,455	-0,064
	$\Delta_{T1-T3}$	5,6 ± 8,9	(426)	6,3 ± 9,6	(164)	0,377	-0,081
<b>VR-12, MCS</b> (Norm: 50)	$\Delta_{T1-T2}$	2,2 ± 10,3	(475)	1,8 ± 9,5	(192)	0,633	0,041
	$\Delta_{T1-T3}$	2,9 ± 10,8	(426)	1,1 ± 11,1	(164)	0,075	0,164
Hinweis: Zur besseren Interpretierbarkeit erhielten in obigem Tabellenabschnitt „verbesserte“ Mittelwertdifferenzen ein positives Vorzeichen – unabhängig von der eigentlichen Polung des jeweiligen Instrumentes.							
<b>Erfolg Behandlung (Patientensicht)</b>	<b>T2</b>	3,1 ± 1,1	(439)	3,1 ± 1,2	(178)	0,688	-0,036
(Einzelitem: 1=sehr gut, 5=schlecht)	<b>T3</b>	2,9 ± 1,1	(402)	3,0 ± 1,2	(150)	0,357	-0,088

IG: Interventionsgruppe, KG: Kontrollgruppe, PCS (VR-12): Physical Component Summary, MCS (VR-12): Mental Component Summary, DASS: Depression Anxiety Stress Scale, MW: Mittelwert, SD: Standardabweichung.

Tab. 17b: Veränderung der zusätzlich ausgewerteten sekundären Zielgrößen nach 3 und 6 Monaten (III Available Cases)

Instrument	Messung	IG		KG		p-Wert t-test	Cohens d
		MW ± SD	n	MW ± SD	n		
<b>Schmerzbeschreibungsliste (SBL)</b> (0-12)	$\Delta_{T1-T2}$	1,4 ± 3,3	(416)	1,3 ± 3,3	(168)	0,717	0,033
	$\Delta_{T1-T3}$	1,6 ± 3,4	(375)	1,4 ± 3,6	(147)	0,498	0,066
<b>Habituelles Wohlbefinden (MFHW)</b> (0-35)	$\Delta_{T1-T2}$	4,1 ± 8,5	(481)	3,7 ± 8,3	(195)	0,582	0,047
	$\Delta_{T1-T3}$	4,7 ± 8,6	(434)	3,8 ± 9,5	(164)	0,240	0,111
<b>Pain Catastrophizing Scale, Hilflösigg.</b> (0-20)	$\Delta_{T1-T2}$	2,3 ± 4,3	(477)	2,3 ± 4,3	(188)	0,873	-0,014
	$\Delta_{T1-T3}$	2,7 ± 4,6	(430)	2,3 ± 4,4	(159)	0,378	0,082
<b>Pain Catastrophizing Scale, Verstärk.</b> (0-12)	$\Delta_{T1-T2}$	1,1 ± 2,5	(472)	1,1 ± 2,6	(187)	0,919	0,009
	$\Delta_{T1-T3}$	1,3 ± 2,6	(428)	1,0 ± 2,6	(156)	0,168	0,129
<b>Pain Catastrophizing Scale, Grübeln</b> (0-20)	$\Delta_{T1-T2}$	2,7 ± 4,7	(469)	2,5 ± 5,0	(184)	0,491	0,060
	$\Delta_{T1-T3}$	2,9 ± 5,0	(422)	2,4 ± 4,8	(156)	0,243	0,109
<b>Pain Catastrophizing Scale, gesamt</b> (0-52)	$\Delta_{T1-T2}$	6,1 ± 10,0	(475)	5,9 ± 10,1	(188)	0,764	0,026
	$\Delta_{T1-T3}$	6,9 ± 10,7	(429)	5,6 ± 10,1	(159)	0,201	0,119
Hinweis: Zur besseren Interpretierbarkeit erhielten in obigem Tabellenabschnitt „verbesserte“ Mittelwertdifferenzen ein positives Vorzeichen – unabhängig von der eigentlichen Polung des jeweiligen Instrumentes.							
<b>Patientenzufriedenheit (ZUF-8)</b>	<b>T2</b>	23,5 ± 5,8	(418)	23,4 ± 5,5	(181)	0,866	0,015
(8=geringe, 32=hohe Zufriedenheit)	<b>T3</b>	23,8 ± 6,0	(380)	23,4 ± 6,1	(154)	0,552	0,057

IG: Interventionsgruppe, KG: Kontrollgruppe, MW: Mittelwert, SD: Standardabweichung.

## 3.2 Evaluationsansatz 2

### 3.2.1 Kosten für Interventions- und Kontrollgruppe

**Tab. 18** bietet eine Übersicht über die im Rahmen der Sekundärdatenanalyse abgerechneten durchschnittlichen Leistungsdaten der hierfür eingeschlossenen PAIN2020-Studienteilnehmer sowohl für den 6-monatigen Vor- als auch den 6-monatigen Nachbeobachtungszeitraum, ausgehend vom Assessment-Termin als Indexereignis.

Im Hinblick auf die Auswertungsvariante der *Gesamtkosten* zeigte sich, dass die ermittelten Kosten pro Studienteilnehmer – mit Ausnahme der stationären Rehaleistungen – im Nachbeobachtungszeitraum in beiden Gruppen grundsätzlich höher ausfielen als bei der Vorbeobachtung. In der Summe ergab sich für die direkten medizinischen Kosten in der KG pro Studienteilnehmer ein etwas niedrigerer Betrag sowie Kostenanstieg, auch wenn diese Unterschiede zur IG statistisch nicht relevant waren. Die jeweilige Höhe des Kostenanstiegs

unterschied sich zum Teil jedoch zwischen den Gruppen und einzelnen Leistungsbereichen. So erhöhten sich die mittleren Kosten in der ambulanten vertragsärztlichen Versorgung, bei den Arzneimitteln sowie den Heil- und Hilfsmitteln in der IG weniger stark, als dies in der KG der Fall war (jeweils  $p < 0,05$ ). Für die ambulanten Krankenhausleistungen ergab sich ein umgekehrtes Bild zugunsten der KG, wobei zu bedenken ist, dass sich in diesem Leistungsbereich im Nachbeobachtungszeitraum die *Programmkosten von PAIN2020* niederschlugen. Während die Screeningpauschale (35 €) in beiden Gruppen abgerechnet wurde, fielen Kosten für das interdisziplinäre multimodale Assessment (690 €) sowie die edukative (62,50 €) und die begleitende (140 € pro Termin) interdisziplinäre multimodale Schmerztherapie nur für die IG an. So sind in den in **Tab. 18** ausgewiesenen ambulanten Krankenhausleistungen in der IG pro Studienteilnehmer 368,95 € und in der KG 12,71 € an Programmkosten für PAIN2020 enthalten (nicht tabellarisch dargestellt).

Ebenso wie in der zuvor beschriebenen Auswertungsvariante entfiel auch bei Betrachtung der *Schmerzbezogenen Kosten* pro Kopf die betragsmäßig größte Kostenposition im Nachbeobachtungszeitraum auf den Bereich der stationären Krankenhausleistungen. Mehr als die Hälfte des Anstiegs der direkten medizinischen Kosten pro Studienteilnehmer war auf den stationären Krankenhaussektor zurückzuführen. Beim Gruppenvergleich zeigte sich, dass für die IG in vielen Leistungsbereichen statistisch bedeutsame, geringere Kostenanstiege zwischen den beiden Beobachtungszeiträumen zu verzeichnen waren. Dies führte trotz Berücksichtigung der Programmkosten dazu, dass die mittleren direkten medizinischen Kosten der IG in Summe diejenigen Aufwendungen der KG nach dem Assessment nicht überstiegen. Ergänzend hierzu sind darüber hinaus die Krankengeldzahlungen und AU-Tage anzuführen, deren Mittelwerte in der IG ebenfalls unterhalb den Werten der KG lagen.

*Hinweis: Ermittelte Kosten basieren auf Abrechnungsdaten!*

In der Literatur finden sich vielfach Hinweise auf den Abrechnungscharakter der Routinedaten der gesetzlichen Krankenversicherung (u.a. Zeidler et al. 2013, Swart et al. 2015). Vor diesem Hintergrund sei darauf hingewiesen, dass es insbesondere im Rahmen der Diagnostik auch dazu kommen konnte, dass nicht durch das Expertenteam als schmerzbezogen definierte ICD-Diagnosen (z.B. eine L40.0 („Psoriasis vulgaris“), die sich im weiteren Verlauf zu einer L40.5 konkretisiert („Psoriasis-Arthropathie“)) mit PAIN2020-spezifischen EBM-Ziffern abgerechnet wurden und es hierdurch in der IG zu etwas höheren durchschnittlichen Programmkosten im direkten Vergleich zu den schmerzbezogenen Pro-Kopf-Kosten für ambulante Krankenhausleistungen kam (369 € > 355 €).

Tab. 18 : Kosten 6 Monate vor und 6 Monate nach dem Assessment (Sekundärdatenanalyse mit Abrechnungsdaten der BARMER Krankenkasse)

Leistungsbereich	<b>Gesamtkosten in Euro</b>				<b>Schmerzbezogene Kosten in Euro</b>			
	IG	KG	p-Wert	R <sup>2</sup>	IG	KG	p-Wert	R <sup>2</sup>
	(n = 531)	(n = 212)			(n = 531)	(n = 212)		
	MW ± SD	MW ± SD	U-Test		MW ± SD	MW ± SD	U-Test	
<b>AMB</b> Vorbeobachtung	527 ± 520	559 ± 513	0,277	0,002	260 ± 337	290 ± 273	0,022	0,007
Nachbeobachtung	552 ± 626	756 ± 614	<0,001	0,047	293 ± 356	479 ± 408	<0,001	0,082
<b>Δ<sub>vor-nach</sub></b>	<b>25 ± 555</b>	<b>197 ± 569</b>	<b>&lt;0,001</b>	<b>0,034</b>	<b>33 ± 345</b>	<b>189 ± 370</b>	<b>&lt;0,001</b>	<b>0,058</b>
<b>KHAMB</b> Vorbeobachtung	27 ± 115	35 ± 122	0,332	0,001	6 ± 35	4 ± 29	0,336	0,001
Nachbeobachtung	407 ± 552	64 ± 140	<0,001	0,041	355 ± 528	43 ± 95	<0,001	0,024
<b>Δ<sub>vor-nach</sub></b>	<b>380 ± 534</b>	<b>28 ± 154</b>	<b>&lt;0,001</b>	<b>0,059</b>	<b>349 ± 520</b>	<b>39 ± 100</b>	<b>&lt;0,001</b>	<b>0,028</b>
<b>KH</b> Vorbeobachtung	299 ± 1.583	509 ± 1.967	0,286	0,002	84 ± 538	60 ± 652	0,174	0,002
Nachbeobachtung	805 ± 2.318	835 ± 2.143	0,208	0,002	519 ± 1.771	629 ± 1.836	0,036	0,006
<b>Δ<sub>vor-nach</sub></b>	<b>506 ± 2.698</b>	<b>326 ± 2.954</b>	<b>0,894</b>	<b>0,000</b>	<b>434 ± 1.847</b>	<b>569 ± 1.968</b>	<b>0,020</b>	<b>0,007</b>
<b>AM</b> Vorbeobachtung	296 ± 2.060	182 ± 621	0,656	0,000	45 ± 301	34 ± 63	0,218	0,002
Nachbeobachtung	302 ± 1.803	235 ± 805	0,200	0,002	35 ± 86	55 ± 112	<0,001	0,021
<b>Δ<sub>vor-nach</sub></b>	<b>7 ± 420</b>	<b>52 ± 237</b>	<b>0,032</b>	<b>0,006</b>	<b>-10 ± 303</b>	<b>20 ± 98</b>	<b>0,015</b>	<b>0,008</b>
<b>HEILHILF</b> Vorbeobachtung	208 ± 413	213 ± 274	0,205	0,002	132 ± 175	159 ± 187	0,062	0,005
Nachbeobachtung	250 ± 476	331 ± 390	<0,001	0,022	160 ± 221	248 ± 255	<0,001	0,032
<b>Δ<sub>vor-nach</sub></b>	<b>42 ± 432</b>	<b>118 ± 356</b>	<b>0,003</b>	<b>0,012</b>	<b>28 ± 237</b>	<b>88 ± 280</b>	<b>0,002</b>	<b>0,013</b>
<b>REHA</b> Vorbeobachtung	22 ± 263	29 ± 246	0,406	0,001	8 ± 137	10 ± 143	0,856	0,000
Nachbeobachtung	7 ± 117	19 ± 271	0,851	0,000	4 ± 85	0 ± 0	0,527	0,001
<b>Δ<sub>vor-nach</sub></b>	<b>-14 ± 288</b>	<b>-11 ± 367</b>	<b>0,554</b>	<b>0,000</b>	<b>-5 ± 162</b>	<b>-10 ± 143</b>	<b>0,637</b>	<b>0,000</b>
<b>Summe</b> Vorbeobachtung	1.378 ± 2.872	1.529 ± 2.486	0,294	0,001	536 ± 903	557 ± 859	0,018	0,008
<b>direkte medizini-</b> Nachbeobachtung	2.323 ± 3.223	2.239 ± 2.790	0,957	0,000	1.366 ± 1.910	1.454 ± 1.921	0,321	0,001
<b>nische Kosten</b> <b>Δ<sub>vor-nach</sub></b>	<b>945 ± 2.896</b>	<b>710 ± 3.162</b>	<b>0,468</b>	<b>0,001</b>	<b>830 ± 2.020</b>	<b>896 ± 1.967</b>	<b>0,878</b>	<b>0,000</b>
<b>KG</b> Vorbeobachtung	140 ± 919	162 ± 974	0,758	0,000	79 ± 591	103 ± 704	0,621	0,000
Nachbeobachtung	232 ± 1.323	282 ± 1.287	0,399	0,001	147 ± 1.082	200 ± 1.057	0,253	0,002
<b>Δ<sub>vor-nach</sub></b>	<b>92 ± 942</b>	<b>120 ± 1.077</b>	<b>0,467</b>	<b>0,001</b>	<b>68 ± 898</b>	<b>97 ± 1.062</b>	<b>0,652</b>	<b>0,000</b>
(in Tagen) <b>AU</b> Vorbeobachtung	9,3 ± 24,8	9,3 ± 26,6	0,388	0,001	4,9 ± 18,1	5,8 ± 24,5	0,283	0,002
Nachbeobachtung	9,5 ± 25,7	9,3 ± 23,3	0,492	0,001	4,9 ± 19,4	6,3 ± 20,1	0,172	0,003
<b>Δ<sub>vor-nach</sub></b>	<b>0,2 ± 23,7</b>	<b>0,0 ± 29,6</b>	<b>0,338</b>	<b>0,001</b>	<b>-0,1 ± 17,3</b>	<b>0,5 ± 26,6</b>	<b>0,006</b>	<b>0,010</b>

MW: Mittelwert, SD: Standardabweichung, IG: Interventionsgruppe, KG: Kontrollgruppe, AMB: Ambulante vertragsärztliche Versorgung, KHAMB: Ambulante Krankenhausleistungen, KH: Stationäre Krankenhausleistungen, AM: Arzneimittel, HEILHILF: Heil- und Hilfsmittel, REHA: Stationäre Rehabilitationsleistungen, AU: Arbeitsunfähigkeitstage, KG: Krankengeld.

Überdies ist zu betonen, dass routinemäßig im W-DWH der BARMER nur ambulante Daten nach § 301 SGB V enthalten sind, im Rahmen des Selektivvertrages von PAIN2020 jedoch auch das Datenverfahren nach § 295 Abs. 1b SGB V Anwendung finden konnte. Bei Abfassung des vorliegenden Evaluationsberichtes ließ sich mithin nicht ausschließen, dass es zu einer Untererfassung von hier ausgewiesenen Programmkosten kam. Die tatsächlich angefallenen Programmkosten ließen sich alleine schon dadurch nicht (auf der Grundlage von Routineabrechnungen) mit hoher Genauigkeit beziffern, da nach Mitteilung des Konsortialführers nicht alle teilnehmenden Zentren alle PAIN2020-Leistungen abrechneten.

Tab. 19: Anteil der Studienteilnehmer mit Kosten größer 0 (Sekundärdatenanalyse)

Leistungs- bereich		Gesamtkosten Kosten > 0			Schmerzbezogene Kosten Kosten > 0		
		IG (N = 531) n %	KG (N = 212) n %	p-Wert Chi <sup>2</sup> -Test	IG (N = 531) n %	KG (N = 212) n %	p-Wert Chi <sup>2</sup> -Test
AMB	Vorb.	522 (98,3)	209 (98,6)	0,785	478 (90,0)	202 (95,3)	0,020
	Nachb.	518 (97,6)	209 (98,6)	0,381	474 (89,3)	203 (95,8)	0,005
KHAMB	Vorb.	51 (9,6)	25 (11,8)	0,374	20 (3,8)	5 (2,4)	0,336
	Nachb.	261 (49,2)	93 (43,9)	0,193	225 (42,4)	85 (40,1)	0,570
KH	Vorb.	52 (9,8)	26 (12,3)	0,321	17 (3,2)	3 (1,4)	0,174
	Nachb.	97 (18,3)	48 (22,6)	0,174	52 (9,8)	33 (15,6)	0,026
AM	Vorb.	439 (82,7)	177 (83,5)	0,790	284 (53,5)	123 (58,0)	0,262
	Nachb.	443 (83,4)	180 (84,9)	0,621	269 (50,7)	139 (65,6)	<0,001
HEILHILF	Vorb.	346 (65,2)	146 (68,9)	0,335	317 (59,7)	139 (65,6)	0,138
	Nachb.	343 (64,6)	159 (75,0)	0,006	320 (60,3)	156 (73,6)	<0,001
REHA	Vorb.	4 (0,8)	3 (1,4)	0,399	2 (0,4)	1 (0,5)	0,854
	Nachb.	2 (0,4)	1 (0,5)	0,854	1 (0,2)	0 (0,0)	0,527
Summe direkte mediz. Kosten	Vorb.	523 (98,5)	209 (98,6)	0,926	498 (93,8)	203 (95,8)	0,294
	Nachb.	524 (98,7)	210 (99,1)	0,673	510 (96,0)	209 (98,6)	0,077
KG	Vorb.	20 (3,8)	9 (4,2)	0,761	14 (2,6)	7 (3,3)	0,621
	Nachb.	29 (5,5)	15 (7,1)	0,400	18 (3,4)	11 (5,2)	0,253
(Tage > 0) AU	Vorb.	191 (36,0)	68 (32,1)	0,314	100 (18,8)	32 (15,1)	0,229
	Nachb.	183 (34,5)	80 (37,7)	0,400	103 (19,4)	50 (23,6)	0,202

IG: Interventionsgruppe, KG: Kontrollgruppe, AMB: Ambulante vertragsärztliche Versorgung, KHAMB: Ambulante Krankenhausleistungen, KH: Stationäre Krankenhausleistungen, AM: Arzneimittel, HEILHILF: Heil- und Hilfsmittel, REHA: Stationäre Rehabilitationsleistungen, AU: Arbeitsunfähigkeitstage, KG: Krankengeld, Vorb.: Vor-, Nachb.: Nachbeobachtung.

Da es generell vorkommen kann, dass die Inanspruchnahme von Leistungen in den verschiedenen Bereichen ganz unterschiedlich ausfällt, wurde für **Tab. 19** eine Dichotomisierung der Stichprobe vorgenommen. Studienteilnehmer wurden hierbei entweder der Gruppe zugerechnet, die in der jeweiligen Rubrik Leistungen in Anspruch genommen hat oder aber derjenigen Gruppe, für die in der jeweiligen Leistungsbibliothek keine Abrechnungsdaten verzeichnet waren. Unter dieser Prämisse ist zum einen zu erkennen, dass bestimmte Leistungen wie stationäre Rehabilitationsaufenthalte (im Rahmen der gesetzlichen Krankenversicherung) oder aber auch Krankengeldzahlungen für die hier betrachtete Stichprobe grundsätzlich nur sehr seltene Ereignisse darstellten. Im Gruppenvergleich zeigte sich zum anderen, dass in der IG im Nachbeobachtungszeitraum anteilmäßig zumeist weniger Studienteilnehmer Leistungen in Anspruch nahmen, als dies in der KG der Fall war.

### 3.2.2 Kosten der gematchten Stichprobe aus dem WDWH der BARMER

Die Stichprobenmerkmale der IG und der gematchten KG können **Tab. 20** entnommen werden. Wie an den deskriptivstatistischen Werten und an den p-Werten zu erkennen ist, zeigt sich auch hier, dass Matchingpartner gefunden werden konnten, die sich in den dargestellten Merkmalen nicht wesentlich von der Gruppe der Interventionsteilnehmer unterscheiden.

Tab. 20: Stichprobenmerkmale nach Matching

Merkmal		Interventions- gruppe		gematchte Kontrollgruppe		p-Wert*
		n 491	% (100)	n 491	% (100)	
<b>Soziodemographie</b>						
Geschlecht	weiblich	322	(65,6)	322	(65,6)	1,000
	männlich	169	(34,4)	169	(34,4)	
Alter	MW ± SD (in Jahren)	55,2 ± 13,0		56,0 ± 12,1		0,317
	Min - Max (in Jahren)	20 - 85		19 - 70		
Familienversichert	nein	465	(94,7)	465	(94,7)	1,000
	ja	26	(5,3)	26	(5,3)	
Bundesland	01 Schleswig-Holstein	15	(3,1)	15	(3,1)	1,000
	02 Hamburg	11	(2,2)	11	(2,2)	
	03 Niedersachsen	47	(9,6)	47	(9,6)	
	04 Bremen	0	(0,0)	0	(0,0)	
	05 Nordrhein-Westfalen	52	(10,6)	52	(10,6)	
	06 Hessen	96	(19,6)	96	(19,6)	
	07 Rheinland-Pfalz	51	(10,4)	51	(10,4)	
	08 Baden-Württemberg	25	(5,1)	25	(5,1)	
	09 Bayern	85	(17,3)	85	(17,3)	
	10 Saarland	2	(0,4)	2	(0,4)	
	11 Berlin	11	(2,2)	11	(2,2)	
	12 Brandenburg	18	(3,7)	18	(3,7)	
	13 Meckl.-Vorpommern	13	(32,7)	13	(32,7)	
	14 Sachsen	40	(8,2)	40	(8,2)	
	15 Sachsen-Anhalt	3	(0,6)	3	(0,6)	
	16 Thüringen	22	(4,5)	22	(4,5)	
<b>Schmerz</b>						
Kopfschmerz	ja	168	(34,2)	173	(35,2)	0,738
	nein	323	(65,8)	318	(64,8)	
Gelenkschmerz	ja	344	(70,1)	349	(71,1)	0,726
	nein	147	(30,0)	142	(29,9)	
Schmerzen Wirbelsäule/ Rücken	ja	463	(94,3)	458	(93,3)	0,509
	nein	28	(65,7)	33	(6,7)	
Schmerzen Weichteil- gewebe	ja	350	(71,3)	356	(72,5)	0,670
	nein	141	(28,7)	135	(27,5)	
Sonstige Schmerzen	ja	458	(93,3)	456	(92,9)	0,802
	nein	33	(6,7)	35	(7,1)	

Tab. 20 (Forts,)

Merkmal	Interventions- gruppe		gematchte Kontrollgruppe		p-Wert*
	n	%	n	%	
	491	(100)	491	(100)	
<b>Leistungsbereich</b>		<b>MW ± SD</b>		<b>MW ± SD</b>	
AMB	in €	264 ± 361		246 ± 346	0,241
KHAMB	in €	14 ± 81		18 ± 82	0,894
KH	in €	65 ± 416		67 ± 503	0,999
AM	in €	33 ± 87		37 ± 132	0,948
HEILHILF	in €	131 ± 167		121 ± 169	0,320
REHA	in €	5 ± 105		9 ± 143	0,566
∑ direkte med. Kosten	in €	512 ± 718		499 ± 803	0,139
AU	in Tagen	4,0 ± 15,7		3,9 ± 14,8	0,965

MW: Mittelwert, SD: Standardabweichung, Min: Minimum, Max: Maximum, AMB: Ambulante vertragsärztliche Versorgung, KHAMB: Ambulante Krankenhausleistungen, KH: Stationäre Krankenhausleistungen, AM: Arzneimittel, HEILHILF: Heil- und Hilfsmittel, REHA: Stationäre Rehabilitationsleistungen, AU: Arbeitsunfähigkeitstage \* Tests auf Gruppenunterschiede für intervallskalierte Merkmale mittels t-Test, für kategoriale Merkmale mittels Chi-Quadrat-Test und für nichtparametrische Merkmale mittels U-Test zum Signifikanzniveau 0,05. Kosten und Au-Tage wurden als parametrisch angenommen.

Die Ergebnisse der Kostenberechnungen werden in **Tab. 21** dargestellt. Für jeden Leistungsbereich werden in den mittleren Spalten der Tabelle die Mittelwerte und Standardabweichungen der in den beiden Folgequartalen dokumentierten Kosten bzw. AU-Tage für die Gruppe der PAIN2020-Interventionspatienten sowie für die aus dem Gesamtbestand der BARMER gematchte Kontrollgruppe gezeigt. Die Veränderungen zwischen den beiden Vorgängerquartalen und den beiden Folgequartalen (Mittelwert und Standardabweichung) sind in der unmittelbar folgenden Zeile zu finden. Als statistischer Test auf Unterschiede zwischen den Kosten in der IG und der KG wurde der nichtparametrische Mann-Whitney U-Test verwendet.

Die rechte Seite der Tabelle zeigt zur Orientierung und zur Einordnung in den Kontext die entsprechenden Ergebnisse für den Vergleich mit der KG aus PAIN2020. Diese sind für die hier betrachteten Fragestellungen allerdings von untergeordneter Bedeutung.

Die Ergebnisse in **Tab. 21** verdeutlichen, dass in nahezu allen Leistungsbereichen in der IG höhere Kosten in den beiden Folgequartalen zu beobachten sind. Dies betrifft insbesondere die Bereiche Krankenhausbehandlung (+463 €), ambulante Behandlung (+85 €), Heil- und Hilfsmittel (+58 €) und ambulante Krankenhausleistungen (+48 €). Bei den Kosten für Arzneimittel, den Kosten für Rehabilitationsleistungen und bei den AU-Tagen finden sich nur geringe Unterschiede zwischen den beiden Gruppen oder deskriptive Unterschiede zugunsten der IG.

Tab. 21: Schmerzbezogene Kosten (in €) in der Nachbeobachtung

Leistungsbereich	Interventions- gruppe (n = 491)	gematchte Kontrollgruppe (n = 491)	Vergleich mit Match p-Wert	Kontrollgruppe PAIN2020 (n = 207)	Vergleich mit Match p-Wert
	MW ± SD	MW ± SD	U-Test	MW ± SD	U-Test
AMB <i>Nachbeobachtung</i>	291 ± 375	206 ± 303	<0,001	387 ± 414	<0,001
<i>Δ<sub>vor-nach</sub></i>	<b>27 ± 400</b>	<b>-40 ± 329</b>	<b>0,006</b>	<b>117 ± 385</b>	<b>&lt;0,001</b>
KHAMB <i>Nachbeobachtung</i>	63 ± 299	15 ± 100	<0,001	18 ± 69	0,013
<i>Δ<sub>vor-nach</sub></i>	<b>49 ± 307</b>	<b>4 ± 76</b>	<b>&lt;0,001</b>	<b>12 ± 77</b>	<b>0,163</b>
KH <i>Nachbeobachtung</i>	550 ± 1.863	87 ± 779	<0,001	543 ± 1.864	<0,001
<i>Δ<sub>vor-nach</sub></i>	<b>486 ± 1.894</b>	<b>20 ± 894</b>	<b>&lt;0,001</b>	<b>461 ± 1.979</b>	<b>&lt;0,001</b>
AM <i>Nachbeobachtung</i>	32 ± 73	37 ± 163	0,400	45 ± 90	0,184
<i>Δ<sub>vor-nach</sub></i>	<b>-1 ± 90</b>	<b>-1 ± 57</b>	<b>0,974</b>	<b>9 ± 82</b>	<b>0,371</b>
HEILHILF <i>Nachbeobachtung</i>	150 ± 220	92 ± 194	<0,001	205 ± 238	<0,001
<i>Δ<sub>vor-nach</sub></i>	<b>19 ± 247</b>	<b>-29 ± 163</b>	<b>&lt;0,001</b>	<b>46 ± 273</b>	<b>&lt;0,001</b>
REHA <i>Nachbeobachtung</i>	9 ± 142	10 ± 149	0,999	0 ± 0	0,359
<i>Δ<sub>vor-nach</sub></i>	<b>4 ± 177</b>	<b>0 ± 207</b>	<b>0,709</b>	<b>-22 ± 225</b>	<b>0,209</b>
Σ direkte <i>Nachbeobachtung</i>	1.095 ± 2.007	446 ± 1.014	<0,001	1.198 ± 1.943	<0,001
med. Kosten <i>Δ<sub>vor-nach</sub></i>	<b>583 ± 2.020</b>	<b>-53 ± 1.132</b>	<b>&lt;0,001</b>	<b>623 ± 1.958</b>	<b>&lt;0,001</b>
(in Tagen) AU <i>Nachbeobachtung</i>	5,0 ± 20,7	2,9 ± 13,0	0,053	4,8 ± 14,1	0,002
<i>Δ<sub>vor-nach</sub></i>	<b>1,0 ± 18,6</b>	<b>-1,0 ± 17,3</b>	<b>0,056</b>	<b>0,2 ± 21,4</b>	<b>&lt;0,001</b>

MW: Mittelwert, SD: Standardabweichung, AMB: Ambulante vertragsärztliche Versorgung, KHAMB: Ambulante Krankenhausleistungen, KH: Stationäre Krankenhausleistungen, AM: Arzneimittel, HEILHILF: Heil- und Hilfsmittel, REHA: Stationäre Rehabilitationsleistungen, AU: Arbeitsunfähigkeitstage.

Ergänzend zu diesen Ergebnissen werden in **Tab. 22** die prozentualen Anteile der Patienten dargestellt, die in den beiden Doppelquartalen (zwei Vorgängerquartale, zwei Folgequartale) in den verschiedenen Leistungsbereichen überhaupt Kosten verursacht haben. Im Vergleich der IG mit der gematchten KG ist zunächst zu erkennen, dass diese Anteile in den Vorgängerquartalen aufgrund der exakten Matching-Bedingungen völlig identisch sind. Im Hinblick auf die Folgequartale zeigen sich hingegen relevante Unterschiede zwischen den Gruppen: Besonders in den Leistungsbereichen der stationären Leistungen (KH), der ambulanten Krankenhausleistungen (KHAMB) und der Heil- und Hilfsmittel (HEILHILF) finden sich in der IG höhere Anteile von Patienten, bei denen Ausgaben aufgetreten sind.

Allein aus diesem Umstand würde es sich erklären lassen, dass in der IG im Vergleich zur gematchten KG bei sonst gleichen Bedingungen höhere Kosten angefallen sind. Wie ein Blick auf die Verteilungen der Ausgaben in den beiden Gruppen verdeutlicht, finden sich auch entsprechende Unterschiede in der Höhe der entstandenen Kosten (**Abb. 12 und 13**). In der gematchten KG (im Diagramm mit WDWH bezeichnet) ist nicht nur ein höherer Anteil von Fällen mit einer fehlenden oder sehr niedrigen Kostenkomponente zu erkennen, die Verteilungsformen zeigen in der PAIN2020-Interventionsgruppe (im Diagramm mit IMA bezeichnet) höhere Anteile an Ausgaben im mittleren oder oberen Bereich des Ausgabenspektrums.



Tab. 22: Anteil an schmerzbezogenen Kosten größer 0 in der Vor- und Nachbeobachtung

Leistungsbereich	Interventions- gruppe (n = 491)		gematchte Kontrollgruppe (n = 491)		Vergleich mit Match p-Wert	Kontrollgruppe PAIN2020 (n = 207)		Vergleich mit Match p-Wert	
	n	%	n	%	Chi <sup>2</sup> -Test	n	%	Chi <sup>2</sup> -Test	
AMB Vorb.	441	(89,8)	441	(89,8)	1,000	188	(90,8)	0,685	
	Nachb.	434	(88,4)	407	(82,9)	0,014	194	(93,7)	<0,001
KHAMB Vorb.	35	(7,1)	35	(7,1)	1,000	18	(8,7)	0,475	
	Nachb.	60	(12,2)	21	(4,3)	<0,001	19	(9,2)	0,011
KH Vorb.	14	(2,9)	14	(2,9)	1,000	5	(2,4)	0,747	
	Nachb.	47	(9,6)	7	(1,4)	<0,001	27	(13,0)	<0,001
AM Vorb.	265	(54,0)	265	(54,0)	1,000	122	(59,0)	0,228	
	Nachb.	222	(45,2)	237	(48,3)	0,337	105	(50,7)	0,553
HEILHILF Vorb.	301	(61,3)	301	(61,3)	1,000	132	(63,8)	0,540	
	Nachb.	282	(57,4)	185	(37,7)	<0,001	141	(68,1)	<0,001
REHA Vorb.	1	(0,2)	2	(0,4)	0,563	2	(50,0)	0,372	
	Nachb.	2	(0,4)	2	(0,4)	1,000	0	(0,0)	0,358
Σ direkte mediz. Kosten	Vorb.	457	(93,1)	457	(93,1)	1,000	194	(93,7)	0,756
	Nachb.	455	(92,7)	426	(86,8)	0,002	197	(95,2)	0,001
(Tage > 0) AU	Vorb.	85	(17,3)	85	(17,3)	1,000	28	(13,5)	0,215
	Nachb.	77	(15,7)	57	(11,6)	0,063	43	(20,8)	0,002

AMB: Ambulante vertragsärztliche Versorgung, KHAMB: Ambulante Krankenhausleistungen, KH: Stationäre Krankenhausleistungen, AM: Arzneimittel, HEILHILF: Heil- und Hilfsmittel, REHA: Stationäre Rehabilitationsleistungen, AU: Arbeitsunfähigkeitstage.

### Fazit zur Matching-Analyse

Die Kernergebnisse der hier vorgestellten Matching-Analyse können in vier Punkten zusammengefasst werden:

- 1) Nach Durchführung eines Propensity-Score-Matching, wie es in seinen Grundzügen im Evaluationskonzept definiert wurde, konnte noch keine ausreichende Vergleichbarkeit der Gruppe der PAIN2020-Interventionspatienten und einer gematchten Kontrollgruppe aus dem Gesamtbestand der BARMER-Versicherten erzielt werden.
- 2) Erst mit der zusätzlichen Berücksichtigung weiterer Matching-Bedingungen, insbesondere des exakten Matchings auf der Basis von gruppierten Kostendaten ergab sich ein zufriedenstellendes Matching-Ergebnis.
- 3) Die Analyse der Kosten in den beiden Folgequartalen zeigte, dass besonders in vier Leistungsbereichen (ambulante Leistungen, ambulante Leistungen im Krankenhaus, stationäre Leistungen und Ausgaben für Heil- und Hilfsmittel) höhere Kosten in der PAIN2020-Interventionsgruppe zu verzeichnen waren.
- 4) Diese Ergebnisse präzisieren und ergänzen die Ergebnisse der Routinedatenanalyse, wie sie anhand der Daten der Interventions- und der Kontrollgruppe erzielt und im Evaluationsbericht beschrieben wurden. Im Unterschied zu den in Kapitel 3.2.1 dargestellten Ergebnissen zeigt sich beim Vergleich der Interventions- mit der gematchten Kontrollgruppe jedoch ein deutlicherer Vorteil zugunsten der Kontrollgruppe. Es ist zu empfehlen, die hier berichteten Analyseverfahren mit einem längeren Zeithorizont zu wiederholen, um mögliche längerfristige Effekte der Intervention erfassen zu können.

Abb. 12: Histogramm der Kostenverteilung in den Bereichen ambulante Leistungen (AMB) und ambulante Krankenhausleistungen (KHAMB)

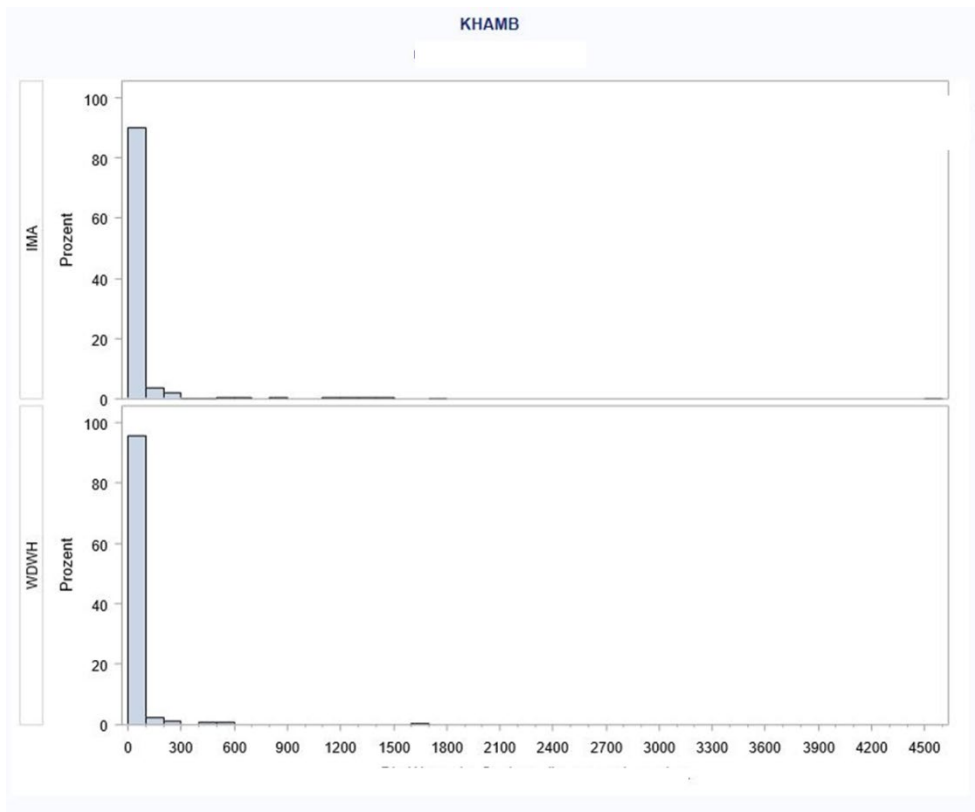
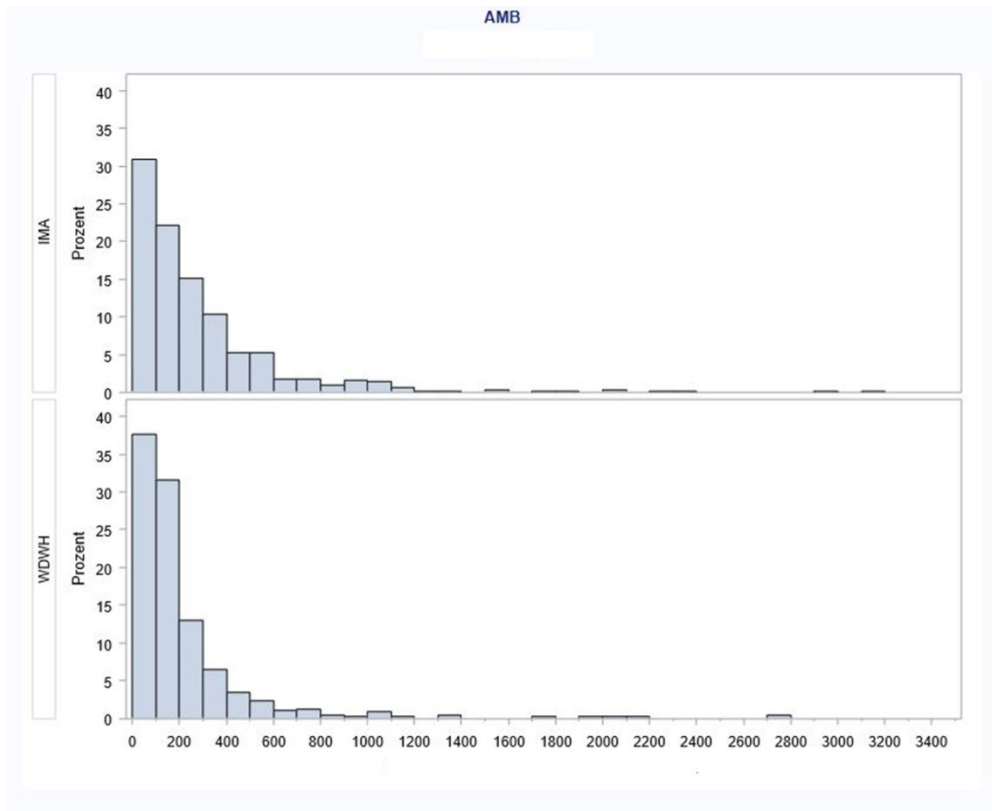
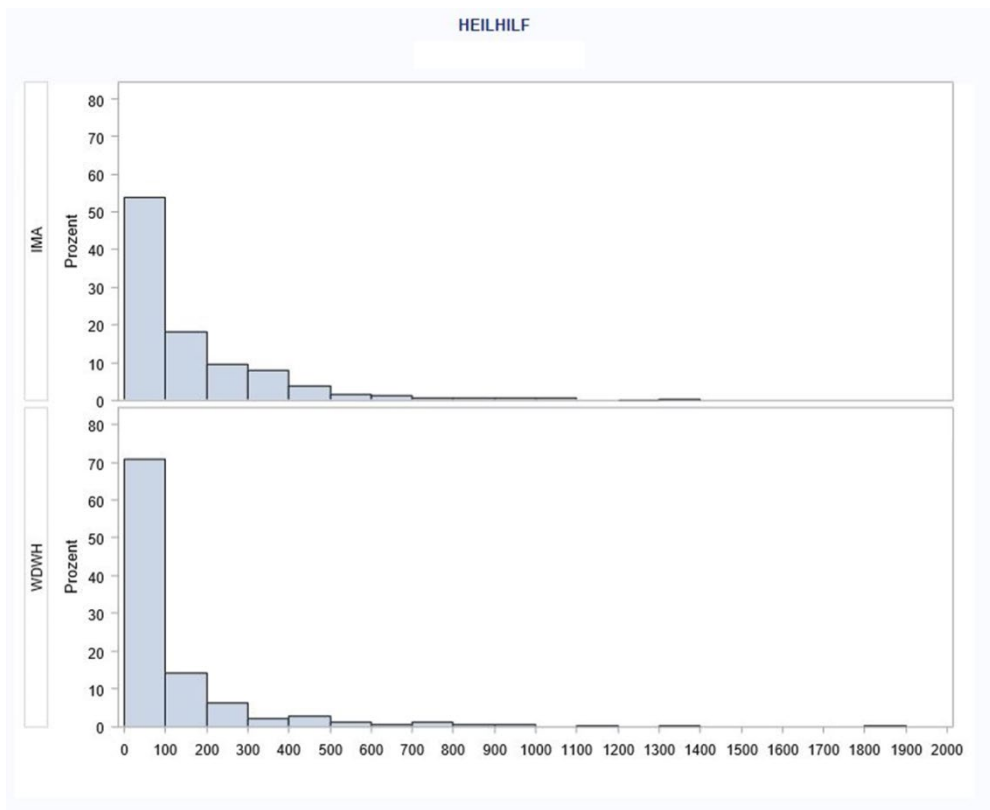
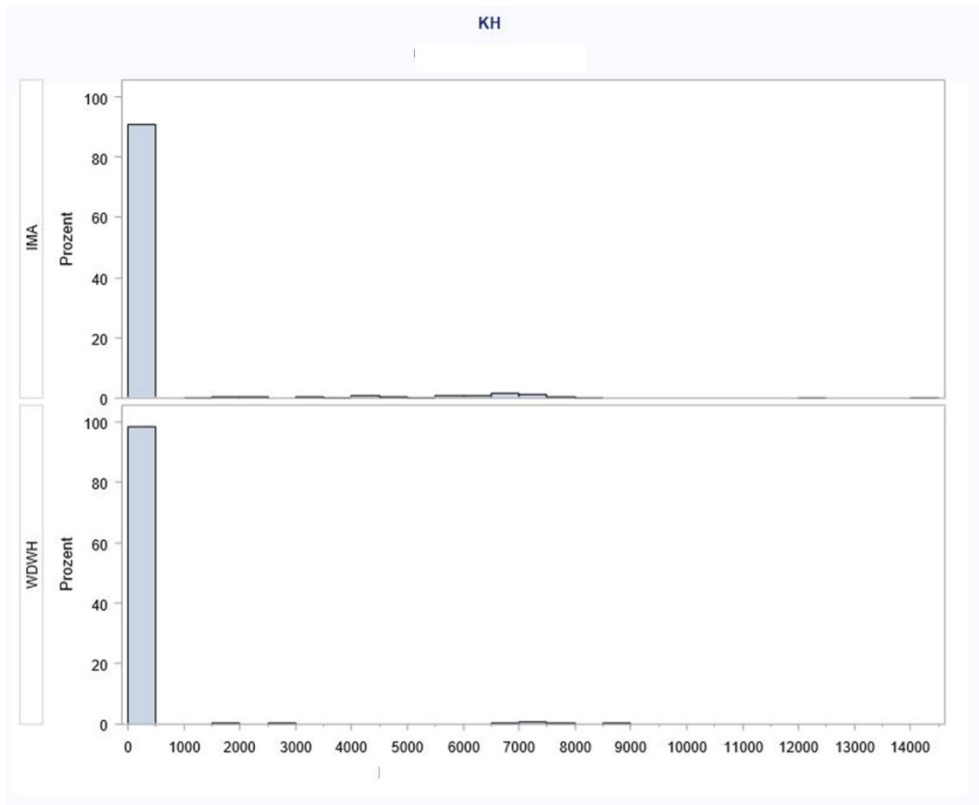


Abb. 13: Histogramm der Kostenverteilung in den Bereichen stationäre Leistungen (KH) sowie Heil- und Hilfsmittel (HEILHILF)



#### 4. Schlussfolgerungen und Empfehlungen des Evaluators

Zum Gesamtverlauf der Evaluation ist zunächst festzuhalten, dass die Evaluationsstudie methodisch ohne relevante Abweichungen von der im Evaluationskonzept (in der überarbeiteten Fassung vom Juli 2020) dargestellten Vorgehensweise durchgeführt werden konnte. Dies betrifft insbesondere die Verwendung aller im Evaluationskonzept vorgesehenen Messinstrumente, das geplante Schema der Messzeitpunkte bei der Erhebung der Primärdaten, die Aufbereitung und Analyse der Sekundärdaten und die Grundkonzepte bei der Datenanalyse.

Eine wesentliche Abweichung von der ursprünglichen Planung ist allerdings im Bereich der Patientenrekrutierung und der erreichten Fallzahlen zu verzeichnen. Aufgrund verschiedener Problemlagen, die sich speziell im Kontext der Covid-19 Pandemie ergaben, konnten nur rund 40% der geplanten Fälle in die Analyse eingeschlossen werden. Obwohl sich die Drop-Out-Rate über die Messzeitpunkte in einem durchaus wesentlichen Maße günstiger als erwartet gestaltete, ist davon auszugehen, dass mit den letztendlich für die Analyse verfügbaren Daten ein entscheidender Powerverlust verbunden war. Dies schränkte die Analysemöglichkeiten ein, auch wenn durch die Randomisierung bedeutsame Baseline-Unterschiede zwischen den Gruppen vermieden werden konnten und sich im weiteren Erhebungsverlauf keine Selektionseffekte manifestierten.

Vor diesem Hintergrund ist als Hauptergebnis der Evaluationsstudie herauszustellen, dass für die beiden primären Zielgrößen, Schmerzintensität und schmerzbedingte Funktionseinschränkung, im Zeitverlauf keine praktisch bedeutsamen und statistisch signifikanten Unterschiede zwischen der Interventions- und der Kontrollgruppe beobachtet wurden. Dieses Ergebnis war übereinstimmend auch in unterschiedlichen Auswertungsvarianten zu beobachten.

Bei den sekundären Zielgrößen und in der Analyse der Routinedaten war zumindest deskriptiv ein Trend zugunsten der Interventionsgruppe erkennbar. Dieser Trend zeigte sich überwiegend im Bereich der psychomentalen Zielgrößen, wie Angst, Depression, Katastrophisieren und dem mentalen Summenscore (MCS) des VR-12. Die Effektgrößen für die Unterschiede zwischen der Interventions- und der Kontrollgruppe erreichten in einigen Fällen Werte über 0,10. Werte in dieser Größenordnung waren in der ursprünglichen Fallzahlplanung für die primären Zielgrößen angenommen worden.

Es sind drei essentielle Schlussfolgerungen und Empfehlungen, die sich aus dem Verlauf und den Ergebnissen der Evaluationsstudie ableiten lassen:

1. Bei einer Weiterführung oder Übernahme der in PAIN2020 entwickelten Interventionselemente in der medizinischen Versorgung sollten zumindest die wichtigsten Messparameter der primären und sekundären Zielgrößen weiterhin systematisch dokumentiert werden. Dadurch könnte der gesundheitliche Verlauf auf der Basis einer größeren Fallzahl umfassender, als es im Rahmen der Evaluationsstudie möglich war, beschrieben werden. Den Messparametern aus dem Bereich der psychomentalen Zielgrößen sollte dabei besonderes Gewicht beigemessen werden.
2. Ein Monitoring der im Behandlungsverlauf entstehenden Kosten und der sozialmedizinischen Effekte speziell auf die Arbeitsunfähigkeit bei Erwerbstätigen sollte unter Verwendung von Routinedaten der Krankenversicherung ebenfalls erfolgen. Entscheidend wäre hier die Beantwortung der Frage, wie sich die Kostensituation auch außerhalb des Kontextes einer Evaluationsstudie darstellt.
3. Soweit möglich, sollte der Zeithorizont bei der Erhebung und Analyse der patientennahen und der ökonomischen Zielgrößen erweitert werden. Es ist plausibel anzunehmen, dass der in der vorliegenden Evaluationsstudie verwendete Zeithorizont von 6 Monaten zu kurz ist, um ein Gesamtbild über den Verlauf der

gesundheitlichen und ökonomischen Effekte liefern zu können. Hier wäre es vermutlich sinnvoll, den zeitlichen Verlauf über eine Periode von mindestens zwei Jahren zu beobachten.

## 5. Literaturverzeichnis

- Buchholz, I., Kohlmann, T., Buchholz, M. (2017). Vergleichende Untersuchung der psychometrischen Eigenschaften des SF-36/SF-12 vs. VR-36/VR-12. Abschlussbericht. [https://www.reha-vffr.de/images/vffrpdf/projekte/2017/VR-Abschlussbericht\\_vffr205.pdf](https://www.reha-vffr.de/images/vffrpdf/projekte/2017/VR-Abschlussbericht_vffr205.pdf). (Stand: 02.11.2022).
- Browne, W.J., Lahi, M.G., Parker, R.M.A. (2009). A Guide to Sample Size Calculations for Random Effect Models via Simulation and the MLPowSim Software Package. University of Bristol. <https://seis.bristol.ac.uk/~frwjb/esrc/MLPOWSIMmanual.pdf> (Stand 31.10.2022).
- Bundesärztekammer (BÄK), Kassenärztliche Bundesvereinigung (KBV), Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) (2017). Nationale VersorgungsLeitlinie Kreuzschmerz. <http://www.kreuzschmerz.versorgungsleitlinien.de> (Stand 27.10.2022).
- Donabedian, A. (1988). The quality of care. How can it be assessed? JAMA 260: 1743-1748.
- Greiner, W., Schöffski, O. (2012). Grundprinzipien einer Wirtschaftlichkeitsuntersuchung. In: Schöffski, O., Graf von der Schulenburg, J.M. Gesundheitsökonomische Evaluationen, 4. Auflage, Heidelberg u.a.O.
- Fritz, C. O., Morris, P. E., Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. Journal of Experimental Psychology: General 141(1): 2-18.
- Häuser, W., Schmutzer, G., Henningsen, P., & Brähler, E. (2014). Chronische Schmerzen, Schmerzkrankheit und Zufriedenheit der Betroffenen mit der Schmerzbehandlung in Deutschland. Schmerz 28(5): 483-492.
- Herda, C., Scharfenstein, A., Basler, H.D. (1998). Marburger Fragebogen zum habituellen Wohlbefinden, Marburg.
- Grobe, T.G., Steinmann, S., Szecsenyi, J. (2016). Barmer GEK Arztreport 2016. <https://www.barmer.de/resource/blob/1026850/41528a9e5704bb8d47e25e00707af4ba/barmer-gek-arztreport-2016-data.pdf> (Stand: 26.10.2022).
- Kazis, L.E., Miller, D.R., Skinner, K.M. et al. (2004). Patient reported measures of health: the veterans health study. J Ambul Care Manag 27:70-83.
- Korb, J., Pfingsten, M. (2003). Der Deutsche Schmerzfragebogen – Implementierte Psychometrie. Schmerz 17: S47.
- Lang, E., Eisele, R., Bickel, A., Winter, E., Schlabeck, M., Kastner, S., Sittl, R., Lievbig, K., Martus, P., Neundörfer, B. (1999). Strukturqualität in der ambulanten Versorgung chronisch schmerzkranker Patienten. Schmerz 13(2): 102-112.
- Lecher, S., Satzinger, W., Trojan, A., Koch, U. (2002). Patientenorientierung durch Patientenbefragungen als ein Qualitätsmerkmal der Krankenversorgung. Bundesgesundheitsbl – Gesundheitsforsch – Gesundheitsschutz 2002; 45: 3–12
- Lovibond, P. F. & Lovibond, S. H. (1995). The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behavioral Research and Therapy 33: 335-343.

- Marschall, U., L'hoest, H. (2013). Schnellwirksame Fentanyl – erhöhtes Suchtpotenzial durch Off-label-use? Untersuchungen einer neueren Stoffklasse auf Basis von Krankenkassendaten. In: Repschläger, U., Schulte, C., Osterkamp, N., Gesundheitswesen aktuell. Wuppertal: BARMER GEK: 270-88.
- Meyer, K., Sprott, H., Mannion, A.F. (2008). Cross-cultural adaptation, reliability, and validity of the German version of the Pain Catastrophizing Scale. *J Psychosom Res* 64(5):469-478.
- Neubauer, S., Zeidler, J., Lange, A., Graf von der Schulenburg, J.M. (2014). Grundlagen und Methoden von GKV-Routinedatenstudien, Diskussionspapier Nr. 534, Hannover. [https://diskussionspapiere.wiwi.uni-hannover.de/pdf\\_bib/dp-534.pdf](https://diskussionspapiere.wiwi.uni-hannover.de/pdf_bib/dp-534.pdf) (Stand: 07.11.2022).
- Nilges, P., Essau, H. (2015). Die Depressions-Angst-Stress-Skalen: Der DASS – ein Screeningverfahren nicht nur für Schmerzpatienten. *Schmerz* 29: 649-657.
- O’Keeffe, M., Purtill, H., Kennedy, N. et al (2016). Comparative effectiveness of conservative interventions for nonspecific chronic spinal pain: Physical, behavioral/psychologically informed, or combined? A systematic review and metaanalysis. *Journal of Pain* 17: 755-774.
- Renker, E., Schlüter, J., Neubauer, E., & Schiltenswolf, M. (2009). Therapy of patients with chronic back pain. Prescriptions--contentment--outcome. *Schmerz* 23(3): 284-291.
- Sachverständigenrat Gesundheit (2001). Bedarfsgerechtigkeit und Wirtschaftlichkeit. Band III: Über-, Unter- und Fehlversorgung. [https://www.svr-gesundheit.de/fileadmin/Gutachten/Gutachten\\_2000\\_2001/Kurzfassung\\_Band3.pdf](https://www.svr-gesundheit.de/fileadmin/Gutachten/Gutachten_2000_2001/Kurzfassung_Band3.pdf) (Stand:27.10.2022).
- Schmidt, J., Lamprecht, F., Wittmann, W.W. (1989). Zufriedenheit mit der stationären Versorgung. Entwicklung eines Fragebogens und erste Validitätsuntersuchungen. *Psychother med Psychol* 39:248-255.
- Sullivan, M.J.L. (2009). The Pain Catastrophizing Scale – User Manual, Montreal, Quebec. [https://aspecthealth.ca/wp-content/uploads/2017/03/PCSMannual\\_English1.pdf](https://aspecthealth.ca/wp-content/uploads/2017/03/PCSMannual_English1.pdf). (Stand: 01.11.2022).
- Swart, E., Ihle, P., Gothe, H., Matusiewicz, D. (2015). Routinedaten im Gesundheitswesen Handbuch Sekundärdatenanalyse: Grundlagen, Methoden und Perspektiven, 2.Auflage, Bern.
- Vickers, A.J., Altman, D.G. (2001). Statistics Notes: Analysing controlled trials with baseline and follow up measurements. *BMJ* 323:1123-1124.
- Viniol, A., Ploner, T., Hickstein, L. et al. (2019). Prescribing practice of pregabalin/gabapentin in pain therapy: an evaluation of German claim data. *BMJ Open* 2019;9:e021535.
- Von Korff, M. (2000), Epidemiologic and Survey Methods: Chronic Pain Assessment. In: Turk, D.C., Melzack, R., Handbook of Pain Assessment, New York.
- Wenig, C.M., Schmidt, C.O., Kohlmann, T., Schweikert, B. (2009), Costs of back Pain in Germany. *Europ J Pain* 13:280-286.
- Zeidler, J., Lange, A., Braun, S. et al. (2013). Die Berechnung indikationsspezifischer Kosten bei GKV-Routinedatenanalysen am Beispiel von ADHS – Ein Methodenvergleich. *Bundesgesundheitsbl* 56:430-438.

Akronym: PAIN2020

Förderkennzeichen: 01NVF17049

## **6. Anhang**

Anhang 1: Fragebogen für Patienten mit länger anhaltenden Schmerzen

Anhang 2: Expertengestützte indikationsspezifische Einschlusskriterien zur Identifizierung schmerzbezogener Kosten und AU-Tage

Anhang 3: Detaillierergebnisse der Analysen mit Zufallskoeffizientenmodellen

## **7. Anlagen**

Nicht zutreffend.



## Anhang 1

Fragebogen für Patienten mit länger anhaltenden Schmerzen



B4

### Fragebogen für Patienten mit länger anhaltenden Schmerzen



Datum: \_\_\_\_\_ Name Patient/in: \_\_\_\_\_

Bitte füllen Sie diesen Fragebogen aus und beantworten jede der Fragen, indem Sie entweder „ja“ oder „nein“ ankreuzen. Die Fragen helfen uns bei der Einschätzung Ihres Krankheitsbildes.

1. Ich befürchte, dass ich eine ernste, bedrohliche Erkrankung habe.  
 ja       nein
2. Wegen der Schmerzen habe ich meine körperlichen Aktivitäten sehr stark reduziert.  
 ja       nein
3. Ich leide neben den Schmerzen auch noch an anderen körperlichen Symptomen (wie z.B. Magen- und Darmverstimmungen, Schwächegefühle, Schwindel oder Hitzewallungen, Luftnot, Herzrasen, Enge in der Brust, Schwitzen, Kälteschauer, Übelkeit, Kribbeln, Taubheitsgefühle, Schlafstörungen etc.).  
 ja       nein
4. Ich war **in den letzten beiden Wochen** meistens traurig oder niedergeschlagen.  
 ja       nein
5. Ich habe mich **in den letzten beiden Wochen** meistens angespannt oder besorgt gefühlt.  
 ja       nein
6. Ich fühle mich in Bezug auf meine Schmerzen sehr stark verunsichert und weiß überhaupt nicht mehr, was richtig oder falsch ist.  
 ja       nein
7. Seit einiger Zeit gibt es bei mir Probleme in der Partnerschaft und/oder in der Familie und/oder dem engeren Freundeskreis.  
 ja       nein
8. Falls Sie in einem Arbeitsverhältnis stehen: Es gibt momentan erhebliche Belastungen/Schwierigkeiten im Zusammenhang mit meinem Arbeitsplatz.  
 ja       nein       Ich bin derzeit nicht berufstätig
9. Auch bei starken Schmerzen neige ich dazu, alles durchzuziehen und meine Aufgaben abzuschließen.  
 ja       nein
10. Ich habe in der letzten Zeit häufiger an schreckliche Ereignisse von **früher** denken müssen oder Träume darüber gehabt (wie z.B. Unfall oder Gewalterfahrung).  
 ja       nein

## Anhang 2

Expertengestützte indikationsspezifische Einschlusskriterien zur Identifizierung schmerzbezogener Kosten und AU-Tage

<b>Schmerzbezogene Diagnosen</b>	
(Klassifikation nach ICD-10 Version 2020 <sup>1</sup> )	
F45.0	<i>Somatisierungsstörung</i>
F45.1	<i>Undifferenzierte Somatisierungsstörung</i>
F45.4	<i>Anhaltende Schmerzstörung</i>
F62.80	<i>Persönlichkeitsstörung durch Schmerz</i>
F68.0	<i>Entwicklung körperlicher Symptome aus psychischen Gründen</i>
G43	<i>Migräne</i>
G44	<i>Sonstige Kopfschmerzsyndrome</i>
G50.0	<i>Trigeminusneuralgie</i>
G50.1	<i>Atypischer Gesichtsschmerz</i>
G50.9	<i>Krankheit des N. trigeminus, n.n.b.</i>
G51.3	<i>Spasmus (hemi)facialis</i>
G51.9	<i>Krankheit des N. facialis, nicht näher bezeichnet</i>
G52.1	<i>Krankheiten des N. glossopharyngeus [IX. Hirnnerv]</i>
G53.0*	<i>Neuralgie nach Zoster (B02.2-Kreuz)</i>
G54.0	<i>Läsionen des Plexus brachialis</i>
G54.1	<i>Läsionen des Plexus lumbosacralis</i>
G54.2	<i>Läsionen der Zervikalwurzeln</i>
G54.3	<i>Läsionen der Thorakalwurzeln</i>
G54.4	<i>Läsionen der Lumbosakralwurzeln</i>
G54.5	<i>Neuralgische Amyotrophie</i>
G54.6	<i>Phantomschmerz</i>
G54.9	<i>Krankheit von Nervenwurzeln und Nervenplexus, n.n.b.</i>
G55*	<i>Kompression von Nervenwurzeln und Nervenplexus bei a.k.K.</i>
G56	<i>Mononeuropathien der oberen Extremität</i>
G57	<i>Mononeuropathien der unteren Extremität</i>
G58	<i>Sonstige Mononeuropathien</i>
G59*	<i>Mononeuropathie bei a.k.K.</i>
G62	<i>Sonstige Polyneuropathien</i>
G63*	<i>Polyneuropathie bei a.k.K.</i>
G72.0	<i>Arzneimittelinduzierte Myopathie</i>
G72.8	<i>Sonstige näher bezeichnete Myopathien</i>
G72.9	<i>Myopathie, n.n.b.</i>
G90.5	<i>Komplexes regionales Schmerzsyndrom, Typ I</i>
G90.6	<i>Komplexes regionales Schmerzsyndrom, Typ II</i>
G90.7	<i>Komplexes regionales Schmerzsyndrom, sonstiger Typ</i>
G95.85	<i>Deafferentierungsschmerz Rückenmark</i>
H57.1	<i>Augenschmerzen</i>
K07.6	<i>Krankheiten des Kiefergelenkes</i>
K14.6	<i>Glossodynie</i>
M03*	<i>Postinfektiöse und reaktive Arthritiden bei a.k.K.</i>
M14*	<i>Arthropathien bei sonstigen a.k.K.</i>
M15	<i>Polyarthrosen</i>
M16	<i>Coxarthrose</i>
M17	<i>Gonarthrose</i>
M18	<i>Rhizarthrose</i>
M19	<i>Sonstige Arthrosen</i>
M22	<i>Krankheiten der Patella</i>
M23	<i>Binnenschädigung des Kniegelenkes</i>

## Anhang 2

Forts.: Expertengestützte indikationsspezifische Einschlusskriterien zur Identifizierung schmerzbezogener Kosten und AU-Tage

<b>Schmerzbezogene Diagnosen (Forts.)</b> (Klassifikation nach ICD-10 Version 2020 <sup>1</sup> )	
M25.5	Gelenkschmerz
M35.3	Polymyalgia rheumatica
M35.7	Hypermobilitäts-Syndrom
M40	Kyphose und Lordose
M41	Skoliose
M42	Osteochondrose der Wirbelsäule
M43.0	Spondylolyse
M43.1	Spondylolisthesis
M43.2	Sonstige Wirbelfusion
M43.8	Sonstige näher bezeichnete Deformitäten der Wirbelsäule und des Rückens
M43.9	Deformität der Wirbelsäule und des Rückens, n.n.b.
M45.0	Spondylitis ankylosans
M47	Spondylose
M48	Sonstige Spondylopathien
M50	Zervikale Bandscheibenschäden
M51	Sonstige Bandscheibenschäden
M53	Sonstige Krankheiten der Wirbelsäule und des Rückens
M54	Rückenschmerzen
M62.6	Muskelzerrung
M62.9	Muskelkrankheit, n.n.b.
M65.2	Tendinitis calcarea
M65.8	Sonstige Synovitis und Tenosynovitis
M65.9	Synovitis und Tenosynovitis, n.n.b.
M70	Bursitiden
M71.4	Bursitis calcarea
M71.5	Sonstige Bursitis, a.n.k.
M71.9	Bursopathie, n.n.b.
M75	Schulterläsionen
M76	Enthesopathien der unteren Extremität
M77	Sonstige Entesioopathien
M79	Sonstige Krankheiten des Weichteilgewebes
M80	Osteoporose mit pathologischer Fraktur
M81	Osteoporose ohne pathologische Fraktur
M89.0	Neurodystrophie (Algodystrophie)
M94.0	Tietze-Syndrom
M94.9	Knorpelkrankheit, n.n.b.
M96.1	Postlaminektomie-Syndrom, a.n.k.
M99	Biomechanische Funktionsstörungen, a.n.k.
N41.1	Chronische Prostatitis
N50.9	Krankheit der männlichen Genitalorgane, n.n.b.
N94	Dysmenorrhoe
R07	Hals- und Brustschmerzen
R10	Bauch- und Beckenschmerzen
R30	Schmerzen beim Wasserlassen
R51	Kopfschmerz
R52	Schmerz, a.n.k.

a.n.k. = anderenorts nicht klassifiziert

a.k.k. = bei anderenorts klassifizierten Krankheiten

n.n.b. = nicht näher bezeichnet

<sup>1</sup> <https://www.dimdi.de/static/de/klassifikationen/icd/icd-10-gm/kode-suche/htmlgm2020/>

## Anhang 2

Forts.: Expertengestützte indikationsspezifische Einschlusskriterien zur Identifizierung schmerzbezogener Kosten und AU-Tage

<b>Schmerzbezogene ATC-Codes</b> (Klassifikation nach amtlicher deutscher Fassung 2020 <sup>1</sup> )	
A03D	SPASMOLYTIKA IN KOMBINATION MIT ANALGETIKA
A03EA	Spasmolytika, Psycholeptika und Analgetika in Kombination
C07AA05	Propranolol <sup>2</sup>
C07AB02	Metoprolol <sup>2</sup>
H02AB	Glucocorticoide
M01A	NICHTSTEROIDALE ANTIPHLOGISTIKA UND ANTIRHEUMATIKA
M02A	TOPISCHE MITTEL GEGEN GELENK- UND MUSKELSCHMERZEN
M03AX	Andere Muskelrelaxanzien, peripher wirkende Mittel
M03BA03	Methocarbamol
M03BX54	Tolperison, Kombinationen
M09AP	Andere pflanzl. Zubereitungen gegen Störungen des Muskel- und Skelettsystems
N01B	LOKALANÄSTHETIKA
N02	ANALGETIKA
N03AF01	Carbamazepin
N03AF02	Oxcarbazepin
N03AG01	Valproinsäure
N03AX12	Gabapentin <sup>3</sup>
N03AX16	Pregabalin
N06AA01	Desipramin
N06AA02	Imipramin
N06AA06	Trimipramin
N06AA10	Nortriptylin
N06AX16	Venlafaxin
N06AX17	Milnacipran
N06AX21	Duloxetin
N06AA09	Amitriptylin
N06AA12	Doxepin
N06AA25	Amitriptylinoxid

<sup>1</sup> [https://www.bfarm.de/DE/Kodiersysteme/Klassifikationen/ATC/Downloads/\\_node.html](https://www.bfarm.de/DE/Kodiersysteme/Klassifikationen/ATC/Downloads/_node.html)

<sup>2</sup> nur, sofern von den Facharztgruppen Anästhesiologie (04), Orthopädie (10), Nervenheilkunde/Neurologie und Psychiatrie (51), Neurologie (53) oder Psychiatrie und Psychotherapie (58) verordnet.

<sup>3</sup> keine Facharztgruppendifferenzierung, da es in der Praxis zu 85% mit Schmerzbezug verordnet wird. (Quelle: Viniol et al. (2019) Prescribing practice of pregabalin/gabapentin in pain therapy: an evaluation of German claim data)

## Anhang 2

Forts.: Expertengestützte indikationsspezifische Einschlusskriterien zur Identifizierung schmerzbezogener Kosten und AU-Tage

<b>Schmerzbezogene Heilmittelindikationen</b> (Klassifikation nach Heilmittelkatalog <sup>1,2</sup> )	
CD	<i>Craniomandibuläre Störungen</i>
CS	<i>Chronifiziertes Schmerzsyndrom (inkl. CSZ)</i>
EX	<i>Erkrankungen der Extremitäten und des Beckens</i>
SB1	<i>Wirbelsäulenerkrankungen</i>
SB2	<i>Störungen nach Trauma, Operation, Verbrennung</i>
SB3	<i>Amputationen, angeborene Fehlbildungen</i>
SB4	<i>Gelenkerkrankungen (kurzzeitiger Behandlungsbedarf)</i>
SB5	<i>Gelenkerkrankungen (längerer Behandlungsbedarf)</i>
SB6	<i>Sympathische Reflexdystrophie, Sudeck'sches Syndrom, CRPS</i>
WS	<i>Wirbelsäulenerkrankungen</i>
EN4	<i>Erkrankungen peripherer Nerven</i>

<sup>1</sup> [https://www.g-ba.de/downloads/17-98-3064/HeilM-RL\\_2019-09-19\\_iK-2020-01-01\\_Heilmittelkatalog.pdf](https://www.g-ba.de/downloads/17-98-3064/HeilM-RL_2019-09-19_iK-2020-01-01_Heilmittelkatalog.pdf)

<sup>2</sup> [https://social.kzvb.de/Zahnaerztliche\\_Heilmittelverordnung/Heilmittelkatalog.pdf](https://social.kzvb.de/Zahnaerztliche_Heilmittelverordnung/Heilmittelkatalog.pdf)

<b>Schmerzbezogene Hilfsmittel</b> (Produktgruppen nach Hilfsmittelverzeichnis des GKV-Spitzenverbands <sup>1</sup> )	
08	<i>Einlagen</i>
09	<i>Elektrostimulationsgeräte<sup>2</sup></i>
05	<i>Bandagen</i>

<sup>1</sup> <https://hilfsmittel.gkv-spitzenverband.de/home>

<sup>2</sup> nur niederfrequente Elektrostimulationsgeräte zur Schmerzbehandlung (Produktgruppe 09.37.01)

### Anhang 3

Detailergebnisse der Analysen mit Zufallskoeffizientenmodellen

#### Primäre Zielgrößen

<b>Schmerzintensität</b>	Regressionskoeffizient	p-Wert	95%-Konfidenzintervall
Konstante	4,23	<0,001	3,91 – 4,55
Ausgangswert T1	0,44	<0,001	0,35 – 0,53
Zeitpunkt T2	0,33	<0,001	0,02 – 0,64
Gruppe IG	-0,09	0,688	-0,47 – 0,29
Zeitpunkt x Gruppe (T2, IG)	0,05	0,781	-0,32 – 0,42

#### **Schmerzbedingte**

#### **Funktionseinschränkung**

	Regressionskoeffizient	p-Wert	95%-Konfidenzintervall
Konstante	3,07	<0,001	2,74 – 3,40
Ausgangswert T1	0,45	<0,001	0,39 – 0,51
Zeitpunkt T2	0,40	<0,001	0,07 – 0,72
Gruppe IG	-0,13	0,666	-0,52 – 0,26
Zeitpunkt x Gruppe (T2, IG)	0,12	0,542	-0,26 – 0,50

#### Sekundäre Zielgrößen

#### **DASS, Angst**

	Regressionskoeffizient	p-Wert	95%-Konfidenzintervall
Konstante	2,45	<0,001	2,09 – 2,81
Ausgangswert T1	0,57	<0,001	0,52 – 0,62
Zeitpunkt T2	-0,08	0,765	-0,43 – 0,28
Gruppe IG	-0,23	0,294	-0,66 – 0,19
Zeitpunkt x Gruppe (T2, IG)	0,09	0,680	-0,33 – 0,50

#### **DASS, Depression**

	Regressionskoeffizient	p-Wert	95%-Konfidenzintervall
Konstante	4,00	<0,001	3,50 – 4,49
Ausgangswert T1	0,59	<0,001	0,54 – 0,64
Zeitpunkt T2	0,07	0,279	-0,41 – 0,55
Gruppe IG	-0,37	0,269	-0,95 – 0,22
Zeitpunkt x Gruppe (T2, IG)	0,17	0,548	-0,40 – 0,74

#### **DASS, Stress**

	Regressionskoeffizient	p-Wert	95%-Konfidenzintervall
Konstante	5,32	<0,001	4,80 – 5,84
Ausgangswert T1	0,55	<0,001	0,50 – 0,60
Zeitpunkt T2	0,19	0,061	-0,36 – 0,74
Gruppe IG	-0,11	0,963	-0,72 – 0,51
Zeitpunkt x Gruppe (T2, IG)	0,23	0,475	-0,41 – 0,88

#### **VR-12, PCS**

	Regressionskoeffizient	p-Wert	95%-Konfidenzintervall
Konstante	41,93	<0,001	40,75 – 43,10
Ausgangswert T1	0,58	<0,001	0,52 – 0,64
Zeitpunkt T2	-1,49	<0,001	-2,61 – -0,38
Gruppe IG	-0,66	0,343	-2,04 – 0,73
Zeitpunkt x Gruppe (T2, IG)	0,17	0,796	-1,14 – 1,50

#### **VR-12, MCS**

	Regressionskoeffizient	p-Wert	95%-Konfidenzintervall
Konstante	46,60	<0,001	45,21 – 47,99
Ausgangswert T1	0,58	<0,001	0,53 – 0,64
Zeitpunkt T2	0,30	0,669	-1,09 – 1,69
Gruppe IG	1,53	0,132	-0,11 – 3,17
Zeitpunkt x Gruppe (T2, IG)	-0,96	0,252	-2,60 – 0,68

### Anhang 3

Forts.: Detailergebnisse der Analysen mit Zufallskoeffizientenmodellen

#### Zusätzlich ausgewertete sekundäre Zielgrößen

<b>Schmerzbeschreibungsliste</b>	Regressionskoeffizient	p-Wert	95%-Konfidenzintervall
Konstante	2,76	<0,001	2,32 – 3,20
Ausgangswert T1	0,41	<0,001	0,35 – 0,47
Zeitpunkt T2	0,17	0,132	-0,28 – 0,63
Gruppe IG	-0,41	0,085	-0,92 – 0,11
Zeitpunkt x Gruppe (T2, IG)	0,07	0,805	-0,47 – 0,61

<b>Habituelles Wohlbefinden</b>	Regressionskoeffizient	p-Wert	95%-Konfidenzintervall
Konstante	21,38	<0,001	20,30 – 22,46
Ausgangswert T1	0,42	<0,001	0,36 – 0,48
Zeitpunkt T2	-0,28	0,212	-1,38 – 0,82
Gruppe IG	0,79	0,220	-0,48 – 2,06
Zeitpunkt x Gruppe (T2, IG)	-0,27	0,681	-1,57 – 1,03

<b>Pain Catastrophizing Scale, Hilflosigkeit</b>	Regressionskoeffizient	p-Wert	95%-Konfidenzintervall
Konstante	5,37	<0,001	4,78 – 5,95
Ausgangswert T1	0,55	<0,001	0,50 – 0,61
Zeitpunkt T2	0,10	0,116	-0,48 – 0,68
Gruppe IG	-0,42	0,402	-1,11 – 0,26
Zeitpunkt x Gruppe (T2, IG)	0,35	0,307	-0,33 – 1,04

<b>Pain Catastrophizing Scale, Verstärkung</b>	Regressionskoeffizient	p-Wert	95%-Konfidenzintervall
Konstante	3,01	<0,001	2,68 – 3,34
Ausgangswert T1	0,49	<0,001	0,43 – 0,55
Zeitpunkt T2	-0,04	0,369	-0,37 – 0,29
Gruppe IG	-0,34	0,193	-0,73 – 0,05
Zeitpunkt x Gruppe (T2, IG)	0,25	0,204	-0,14 – 0,64

<b>Pain Catastrophizing Scale, Grübeln</b>	Regressionskoeffizient	p-Wert	95%-Konfidenzintervall
Konstante	6,04	<0,001	5,40 – 6,68
Ausgangswert T1	0,48	<0,001	0,43 – 0,54
Zeitpunkt T2	-0,03	0,463	-0,68 – 0,63
Gruppe IG	-0,64	0,137	-1,39 – 0,11
Zeitpunkt x Gruppe (T2, IG)	0,34	0,384	-0,43 – 1,11

<b>Pain Catastrophizing Scale, Gesamt</b>	Regressionskoeffizient	p-Wert	95%-Konfidenzintervall
Konstante	14,47	<0,001	13,09 – 15,85
Ausgangswert T1	0,57	<0,001	0,51 – 0,62
Zeitpunkt T2	0,02	0,238	-1,35 – 1,39
Gruppe IG	-1,41	0,171	-3,03 – 0,20
Zeitpunkt x Gruppe (T2, IG)	0,94	0,254	-0,68 – 2,55